Linear Continuous Interior Penalty Finite Element Method for Helmholtz Equation with High Wave Number: One Dimensional Analysis

Erik Burman[1] Haijun Wu[2] Lingxue Zhu[3]
Department of Mathematics, University College London, London, UK-WC1E 6BT, United Kingdom
Department of Mathematics, Nanjing University, Jiangsu, 210093, P.R.China
Department of Mathematics, Jinling Institute of Technology, Jiangsu, 211169, P.R.China

This paper addresses the properties of Continuous Interior Penalty (CIP) finite element solutions for the Helmholtz equation. The h-version of the CIP finite element method with piecewise linear approximation is applied to a one-dimensional model problem. We first show discrete well posedness and convergence results, using the imaginary part of the stabilization operator, for the complex Helmholtz equation. Then we consider a method with real valued penalty parameter and prove an error estimate of the discrete solution in the H^1-norm, as the sum of best approximation error plus a pollution term that is the order of the phase difference. It is proved that the pollution effect can be eliminated by selecting the penalty parameter appropriately. As a result of this analysis, thorough and rigorous understanding of the error behavior throughout the range of convergence is gained. Numerical results are presented that show sharpness of the error estimates and highlight some phenomena of the discrete solution behavior. In particular, we give numerical evidence that the optimal penalty parameter obtained in one-dimensional case also works very well for the CIP-FEM on two-dimensional Cartesian grids.

Keywords: Helmholtz equation, high wave number, pollution, continuous interior penalty finite element methods, error estimates

[1](e.burman@ucl.ac.uk). The work of the first author was partially supported by EPSRC grant no EP/J002313/2.
[2](hjw@nju.edu.cn). The work of the second author was partially supported by the NSF of China grants 11525103 and 91130004.
[3](zlx1987@jit.edu.cn). The work of the third author was partially supported by the National Natural Science Foundation of China grant 11401272 and by the Natural Science Foundation of Jiangsu Province of China grant BK20140105 and by the doctoral scientific research foundation of Jinling Institute of Technology grant jit-6-201413.

Numerical Methods for Partial Differential Equations , 1—32 ()
 CCC

I. INTRODUCTION

The numerical solution of Helmholtz equation using the finite element method (FEM) in the medium to high wave number remains a challenge due to the strong pollution effects that are present in this regime. It is known that when the standard Galerkin method is used a so called scale resolution condition must be satisfied (see [26]) in order to achieve a quasi optimality estimate that is robust in the wave number k. Invertibility of the linear system also holds only under certain conditions on the relation between k and the discretization parameters h and p. This in particular imposes the use of high order finite elements and seems to exclude the possibility of using the simplest choice of piecewise affine elements. In this latter case the standard Galerkin finite element method has to be modified in order to obtain an efficient method. Such modifications often take the form of least squares terms giving additional control of certain residual quantities, either in the element or on element faces. For low order finite elements there are a number of works on stabilized methods, typically using Galerkin least squares approaches and some results on the effect of the stabilization on the dispersion error exist in the one dimensional case, see [21], or for an early example of the use of face based residuals see [27]. Recently a variational multiscale approach using subscales was applied to the Helmholtz equation, leading to a method similar to the one analysed herein [7]. Another possibility is to use discontinuous Galerkin methods and in this framework it has been proven by Feng and Wu [19] that provided a penalty on the jumps of derivatives over element faces is added to the formulation the linear system is always invertible. Similar results were obtained using the continuous interior penalty finite element method in a recent work by Wu [31] and numerical investigations showed that the pollution error could be greatly reduced by choosing the penalty parameter appropriately. For wave-number-explicit error analyses of other methods including spectral methods and discontinuous Petrov-Galerkin methods, we refer to [13, 14, 25, 29, 34].

In the present work we continue the investigations initiated in [31], this time focusing on the one dimensional case and the effect of the penalty operator on the errors in amplitude and phase. Throughout the paper, C is used to denote a generic positive constant which is independent of k, h, f. C may have different values in different occurrences. We also use the shorthand notations $A \lesssim B$ and $B \lesssim A$ for the inequalities $A \leq CB$ and $B \leq CA$. $A \eqsim B$ is for the statement $A \lesssim B$ and $B \lesssim A$. First we will give alternative proofs of some of the results given in [31], showing for methods using a penalty parameter with non-zero imaginary part, that the linear system is always well posed and the following error estimate holds

$$\|(u - u_h)'\| \lesssim (kh + \min(1, k^3h^2)) \|f\| \tag{1.1}$$

for $kh \lesssim 1$ and $k \gtrsim 1$, where $\|\cdot\|$ denotes the L^2-norm. Then we consider the case when the penalty parameter is real and by constructing the discrete Green's function we derive an error estimate where the error is written as the sum of the best approximation error and a term proportional to the phase error, that is

$$\|(u - u_h)'\| \lesssim (kh + \left|k_h^- - k\right|) \|f\|, \quad \text{if } kh \leq 1, \text{ and } 0 < |\gamma| \leq \frac{1}{6}, \tag{1.2}$$

where k_h^- is some discrete wavenumber and γ is the penalty parameter. Compared with the results on linear CIP-FEM in two and three dimensions (see [31, 33, 17]), the above estimate (1.2) for the case of real penalty parameters is unique in two ways. First,

the mesh condition $kh \lesssim 1$ is weaker than the requirement that k^3h^2 is bounded, the latter being the condition required in general dimensions. Secondly, (1.2) says that the pollution error is bounded by the phase difference $O\big(\,|k_h^- - k|\,\big)$, which has never been proved rigorously before. We prove a relation between the phase error and the penalty parameter and show that for a particular range of values for the penalty parameter, under a mild condition on the computational mesh, the phase error is $O(kh)$ and hence the pollution error is eliminated, leading to the pollution-free error estimate

$$\|(u - u_h)'\| \lesssim kh\,\|f\|\,. \tag{1.3}$$

These results are finally verified computationally in several numerical examples. In particular we observe that, when the optimal penalty parameter from the one dimensional analysis is applied to two-dimensional simulations on Cartesian grids, the pollution error of the CIP-FEM is reduced significantly. We remark that, although the analyses of this paper are done in one dimension, they appear to yield information on the choice of the parameter useful also in the higher dimensional case and improves our understanding of the behavior of the CIP-FEM in higher dimensions.

This paper is organized as follows. In Section 2 we study the one-dimensional model problem and introduce the CIP-FEM. Preasymptotic error estimates in H^1- and L^2-norms are derived in Section 3 for any $k > 0$, $h > 0$ and imaginary penalty parameters. In Section 4, we consider the dispersion analysis of the CIP-FE method and obtain the phase error estimates between the wave number k of the continuous problem and the discrete wave number k_h^- for different real penalty parameters. The discrete global system was solved explicitly in Section 5 via the theory of fundamental system and plays a major part in the stability and preasymptotic error analysis. In Section 6, the stability and error estimates are proved directly and we show that the pollution effect can be eliminated by choosing the appropriate penalty parameter. Extensive numerical tests are given in Section 7 to show some phenomena of the discrete solution behavior and verify the theoretical findings, and we come to the conclusion in Section 8.

II. THE MODEL PROBLEM AND ITS DISCRETIZATION

A. The Boundary Value Problem

Let $\Omega = (0,1)$ and let on $\bar{\Omega}$ the boundary value problem (BVP) $Lu = -f$ on be given:

$$u''(x) + k^2 u(x) = -f(x), \quad x \in \Omega \tag{2.1}$$

$$u(0) = 0, \tag{2.2}$$

$$u'(1) - \mathbf{i}ku(1) = 0, \tag{2.3}$$

where, for simplicity, $f(x) \in L^2(\Omega)$ and k is known as the wave number. We assume that $k \gtrsim 1$. Actually, we are interested in wavenumber so large that $k \gg 1$ since we are considering high-frequency problems.

Notation

By $L^2(\Omega) := H^0(\Omega)$, we denote the space of all square-integrable complex-valued functions equipped with the inner product

$$(v, w) := \int_\Omega v(x)\bar{w}(x)\,\mathrm{d}x \text{ and the norm } \|w\| := \sqrt{(w, w)}.$$

We use the notation $H^s(\Omega)$ the Sobolev spaces of (integer) order s in the usual sense. Let $\|\cdot\|_s$ and $|\cdot|_s$ denote the usual full norm and seminorm on H^s, respectively.

Existence and Uniqueness in $H^2(0,1)$

The BVP (2.1)–(2.3) has a unique solution in the space $H^2(0,1)$. For the proof see, e.g., [3]. The existence of the solution is concluded from the following construction.

Inverse Operator

The Green's function of the BVP (2.1)–(2.3) is

$$G(x,s) = \frac{1}{k}\begin{cases} \sin kx\, e^{\mathbf{i}ks}, & 0 \le x \le s, \\ \sin ks\, e^{\mathbf{i}kx}, & s \le x \le 1. \end{cases} \tag{2.4}$$

The solution $u(x)$ of (2.1)–(2.3) exists for all $k > 0$ and can be written as

$$u(x) = \int_0^1 G(x,s) f(s)\,\mathrm{d}s, \tag{2.5}$$

and we have,

$$u'(x) = \int_0^1 H(x,s) f(s)\,\mathrm{d}s \tag{2.6}$$

where $H(x,s) = \begin{cases} \cos kx\, e^{\mathbf{i}ks}, & 0 \le x < s, \\ \mathbf{i}\sin ks\, e^{\mathbf{i}kx}, & s < x \le 1. \end{cases}$

Lemma 2.1. *The BVP (2.1)–(2.3) has a unique solution in $H^2(0,1)$ and for $f \in L^2(0,1)$*

$$\|u\| \le k^{-1}\|f\|, \tag{2.7}$$
$$|u|_1 \le \|f\|, \tag{2.8}$$
$$|u|_2 \le (1+k)\|f\|. \tag{2.9}$$

Proof. The results can be obtained easily by integrating the squares of equations (2.5)–(2.6) and using (2.1) and (2.4). We omitted the details. See also Douglas *et al.* [16]. ∎

Remark 2.1. The aforementioned results are valid also for the adjoint problem (2.1), (2.2) and $u'(1) + \mathbf{i}ku(1) = 0$.

B. The Continuous Interior Penalty Finite Element method

Let $\mathcal{M}_h$ be a uniform mesh on $\bar{\Omega}$ that consists of n sub-intervals $K_j = (x_{j-1}, x_j)$, $1 \le j \le n$, where $x_j = j/n$. Note that $x_j, 1 \le j \le n-1$ are interior nodes and x_0 is the Dirichlet boundary node. The stepsize is $h = 1/n$. For the ease of presentation, we assume that k is constant on Ω.

For any function v, denote by $v_j^+ = v(x_j+)$ and $v_j^- = v(x_j-)$ if the one-sided limits exist. We also define the jump $[v]_j$ of v at the node x_j as

$$[v]_j := v_j^- - v_j^+, \qquad 1 \le j \le n-1.$$

Now we define the "energy" space V and the sesquilinear form $a_h(\cdot,\cdot)$ on $V \times V$ as follows:

$$V := \{v \in H^1(\Omega) \wedge v(0) = 0\} \cap \prod_{K_j \in \mathcal{M}_h} H^2(K_j), \quad j = 1, 2, \cdots, n,$$

$$a_h(u,v) := (u', v') - k^2(u,v) - \mathbf{i}ku(1)\bar{v}(1) + J(u,v) \quad \forall u, v \in V, \tag{2.10}$$

where

$$J(u,v) := \sum_{j=1}^{n-1} \gamma h [u']_j [\bar{v}']_j + \gamma h (u'(1) - \mathbf{i}ku(1))(\bar{v}'(1) - \mathbf{i}k\bar{v}(1)) \tag{2.11}$$

and $\gamma := \gamma_{\mathtt{Re}} + \mathbf{i}\gamma_{\mathtt{Im}}$ is a complex number.

Remark 2.2. (a) The terms in $J(u,v)$ are so-called penalty terms. The penalty parameter in $J(u,v)$ is γ. Observe that if u is the solution of (2.1)-(2.3) then $J(u, v_h) = 0$ for all $v_h \in V_h$ and if z is the adjoint solution discussed in Remark 2.1, then $J(v_h, z) = 0$ for all $v_h \in V_h$.

(b) Penalizing the jumps of normal derivatives was used early by Douglas and Dupont [15] for second order PDEs and by Babuška and Zlámal [6] for fourth order PDEs in the context of C^0 finite element methods, by Baker [8] for fourth order PDEs and by Arnold [2] for second order parabolic PDEs in the context of IPDG methods. More recently it has been proposed and analysed for fourth order PDEs by Hughes et al [18] and for singularly perturbed elliptic or parabolic problems by Burman and co-workers [10, 11, 12].

(c) Notice that we here add a least squares penalty on the boundary condition as well. This enhances the continuity of the sesquilinear form and appears to be necessary for the a priori error estimate proposed below. We remark that this trick has been used to construct stable CIP-FEM for large cavity problems in two dimensions [32].

It is clear that $J(u,v) = 0$ if $u \in H^2(\Omega)$ is the solution of (2.1)–(2.3) and $v \in V$. Therefore,

$$a_h(u,v) = (f,v), \quad \forall v \in V. \tag{2.12}$$

Let V_h be the linear finite element space, that is,

$$V_h := \left\{ v_h \in H^1(\Omega) : \ v_h(0) = 0, v_h|_{K_j} \text{ is a linear polynomial, } j = 1, \cdots, n \right\}.$$

Then our CIP-FEMs are defined as follows: Find $u_h \in V_h$ such that

$$a_h(u_h, v_h) = (f, v_h) \qquad \forall v_h \in V_h. \tag{2.13}$$

We remark that if the parameter $\gamma \equiv 0$, then the above CIP-FEM becomes the standard FEM.

The following semi-norms on the space V are useful for the subsequent analysis:

$$\|v\|_{1,h}^2 := \|v'\|^2 + \sum_{j=1}^{n-1} |\gamma| h \left| [v']_j \right|^2, \tag{2.14}$$

$$\|v\|_{J,k}^2 := |\gamma| \left(\sum_{j=1}^{n-1} h |[v']_j|^2 + h|v'(1)|^2 \right) + k|v(1)|^2, \tag{2.15}$$

$$\|v\|_{J,k,*}^2 := \|v\|_{J,k}^2 + |\gamma|^{-1} \sum_{j=1}^{n} h^{-1} |v(x_j)|^2. \tag{2.16}$$

III. A PRIORI ERROR ESTIMATE FOR THE MODEL PROBLEM

In this section we will use techniques similar to those developed in [10] to derive an a priori error estimate that holds under a very mild condition of $kh \lesssim 1$. We present the analysis in the one dimensional case, but the extension to higher dimensions is straightforward (cf. [31]). The key observations are

1. if the complex component of the penalty parameter is strictly negative (or positive depending on the sign of the boundary condition), the formulation is coercive on the norm $\|\cdot\|_{J,k}$ related to the stabilization;
2. if the L^2-projection is used for interpolation in the analysis, the zeroth order term vanishes and the sesquilinear form $a_h(\cdot,\cdot)$ has enhanced continuity properties.

These two observations lead to an a priori error estimate on the stabilization operator that is optimal in h. An energy norm approach combined with a duality argument is then used to derive an a priori error estimate of the error in the energy norm. To simplify the notation in this section we assume that $\gamma := \mathbf{i}\gamma_{\mathtt{Im}}$. The extension to non-zero real part is straightforward.
Let $\pi_h : L^2(\Omega) \mapsto V_h$ be the standard L^2-projection on V_h. By using the approximation properties of V_h, the trace and inverse inequalities (cf. [9]), it is straightforward to show that

$$\|u - \pi_h u\| + h\|(u - \pi_h u)'\| \lesssim h^2 |u|_2 \tag{3.1}$$

and

$$\|u - \pi_h u\|_{J,k} \lesssim (|\gamma|^{\frac{1}{2}} + (kh)^{\frac{1}{2}})h|u|_2,\ \|u - \pi_h u\|_{J,k,*} \lesssim (|\gamma|^{\frac{1}{2}} + (kh)^{\frac{1}{2}} + |\gamma|^{-\frac{1}{2}})h|u|_2. \tag{3.2}$$

In the following we will assume that $kh \lesssim 1$ and neglect high order contributions in kh in the above approximation estimates. We first prove the continuity of $a_h(\cdot,\cdot)$ on the space orthogonal to V_h. Let

$$V_h^{\perp} := \{v \in V : (v, w_h) = 0, \forall w_h \in V_h\}.$$

Lemma 3.1. *Assume that $|\gamma|\, kh \lesssim 1$. For all $v \in V_h^{\perp}$ and all $w_h \in V_h$ there holds*

$$|a_h(v, w_h)| \lesssim \|v\|_{J,k,*} \|w_h\|_{J,k}$$

and

$$|a_h(v, w_h)| \lesssim \|v\|_{J,k,*} \|w_h - z\|_{J,k}$$

where z denotes the adjoint solution of Remark 2.1.
Proof. The proof follows by observing that

$$a_h(v, w_h) = (v', w_h') - \mathbf{i}kv(1)\overline{w_h}(1) + J(v, w_h).$$

Noting that w_h is piecewise linear and after an integration by parts in the first term in the right hand side we have

$$a_h(v,w_h) = \sum_{j=1}^{n-1} v(x_j)[\overline{w_h}']_j + v(1)(-\mathbf{i}k\overline{w_h}(1) + \overline{w_h}'(1)) + J(v,w_h)$$
$$= \sum_{j=1}^{n-1} v(x_j)[\overline{w_h}' - \overline{z}']_j + v(1)(-\mathbf{i}k(\overline{w_h}(1) - \overline{z}(1)) + (\overline{w_h}'(1) - \overline{z}'(1))) + J(v,w_h - z)$$

where we have used $[\overline{z}']_j = 0$ and the boundary condition $\overline{z}'(1) - \mathbf{i}k\overline{z}(1) = 0$ in the second equality. We conclude by applying the Cauchy-Schwarz inequality and the inequality

$$J(v,w_h - z) \lesssim \|v\|_{J,k}\|w_h - z\|_{J,k}$$

■

For the norm $\|\cdot\|_{J,k}$ we have the following stability estimate.

Lemma 3.2. *Assume that $\gamma = \mathbf{i}\gamma_{\mathtt{Im}}$ with $\gamma_{\mathtt{Im}} < 0$, $|\gamma|\, kh \le 1/2$. Then for all $v_h \in V_h$ there holds*

$$\frac{1}{2}\|v_h\|_{J,k}^2 \le -\mathrm{Im}[a_h(v_h,v_h)]$$

and for u_h solution to (2.13) *then*

$$\frac{1}{2}\|u_h\|_{J,k}^2 \le -\mathrm{Im}[(f,u_h)].$$

Proof. By the definition of $a_h(\cdot,\cdot)$ there holds

$$-\mathrm{Im}[J(v_h,v_h)] + k|v_h(1)|^2 = -\mathrm{Im}[a_h(v_h,v_h)] \tag{3.3}$$

Then observe that for the penalty on the impedance boundary condition we have

$$-\mathrm{Im}\big(\gamma h(v_h'(1) - \mathbf{i}kv_h(1))(\overline{v}_h'(1) - \mathbf{i}k\overline{v}_h(1))\big)$$
$$= |\gamma|\mathrm{Re}\big(h|v_h'(1)|^2 - \mathbf{i}hk\big(v_h(1)\overline{v}_h'(1) + \overline{v}_h(1)v_h'(1)\big) - h|kv_h(1)|^2\big)$$
$$= |\gamma|h|v_h'(1)|^2 - |\gamma|h|kv_h(1)|^2 \ge \frac{1}{2}\big(|\gamma|h|v_h'(1)|^2 - k|v_h(1)|^2\big).$$

The claim follows by using this inequality in (3.3). ■

Remark 3.1. Lemma 3.2 implies existence of a unique discrete solution, since $\|\cdot\|_{J,k}$ is a norm on V_h.

Combining the two previous results with the consistency of the formulation and the regularity estimate (2.9) immediately gives us a convergence estimate for the error in the norm $\|\cdot\|_{J,k}$.

Proposition 3.3. *Let $u \in H^2(\Omega)$ be the solution of* (2.1)–(2.3) *and $u_h \in V_h$ be the solution of* (2.13)*, with γ satisying the assumptions of Lemma 3.2. Then there holds*

$$\|u - u_h\|_{J,k} \lesssim \big(|\gamma|^{\frac{1}{2}} + |\gamma|^{-\frac{1}{2}}\big)kh\|f\|.$$

Proof. Let $u - u_h = \eta - \xi_h$ with $\eta = u - \pi_h u$ and $\xi_h = u_h - \pi_h u$. By the triangle inequality, the error estimate (3.2), and the bound (2.9), it is enough to consider $\|\xi_h\|_{J,k}$. Using Lemma 3.2 followed by the consistency we have

$$\frac{1}{2}\|\xi_h\|_{J,k}^2 \leq -\mathrm{Im}[a_h(\xi_h,\xi_h)] = -\mathrm{Im}[a_h(\eta,\xi_h)] \leq |a_h(\eta,\xi_h)|.$$

We then apply the continuity of Lemma 3.1 to bound the right hand side,

$$\frac{1}{2}\|\xi_h\|_{J,k}^2 \lesssim \|\eta\|_{J,k,*}\|\xi_h\|_{J,k}.$$

Hence,

$$\frac{1}{2}\|\xi_h\|_{J,k} \lesssim \|\eta\|_{J,k,*}, \tag{3.4}$$

then the claim follows by applying once again (3.2) and $(kh)^{\frac{1}{2}} \lesssim |\gamma|^{-\frac{1}{2}}$. ∎

After these preliminary results we are in a position to prove the main result of this section.

Theorem 3.4. *(A priori error estimate)*
Let $u \in H^2(\Omega)$ be the solution of (2.1)–(2.3) and $u_h \in V_h$ the solution of (2.13), with $\gamma_{\mathtt{Im}} < 0$. Then, if h is small such that $kh \lesssim 1$ and $|\gamma|\, kh \leq 1/2$ for all $h > 0$ and $k \gtrsim 1$, there holds

$$\|u - u_h\|_{1,h} \lesssim (|\gamma| + |\gamma|^{-1})\big(kh + \min(1, k^3h^2)\big)\|f\|.$$

Proof. Using once again the decomposition $u - u_h = \eta - \xi_h$, by the estimates (3.1)–(3.2) and Proposition 3.3, we only need to show that

$$\|\xi_h'\| \lesssim (|\gamma| + |\gamma|^{-1})\big(kh + \min(1, k^3h^2)\big)\|f\|. \tag{3.5}$$

To do so, we first prove that

$$\|k\xi_h\| \lesssim (|\gamma| + |\gamma|^{-1})\min(1, k^3h^2)\|f\| \tag{3.6}$$

Consider the adjoint problem, find $z \in H^2(\Omega)$ such that

$$(w', z') - k^2(w, z) - \mathbf{i}kw(1)\bar{z}(1) = (w, \xi_h) \quad \forall w \in V \tag{3.7}$$

and its finite element equivalent, find $z_h \in V_h$ such that

$$a_h(w_h, z_h) = (w_h, u_h - \pi_h u) \quad \forall w_h \in V_h. \tag{3.8}$$

By Lemma 3.2 and Proposition 3.3, z_h exists (cf. Remark 3.1) and satisfies

$$\|z - z_h\|_{J,k} \lesssim (|\gamma|^{\frac{1}{2}} + |\gamma|^{-\frac{1}{2}})kh\|\xi_h\|.$$

Using the consistency of the formulation and the second continuity of Lemma 3.1 followed by the (3.2) we get

$$\begin{aligned}\|\xi_h\|^2 &= a_h(\xi_h, z_h) = a_h(\eta, z_h)\\ &\lesssim \|\eta\|_{J,k,*}\|z - z_h\|_{J,k}\\ &\lesssim (|\gamma| + |\gamma|^{-1})(kh)^2\|f\|\|\xi_h\|.\end{aligned}$$

Therefore,

$$\|k\xi_h\| \lesssim (|\gamma| + |\gamma|^{-1})k^3h^2\|f\|. \tag{3.9}$$

Next we show that $\|k\xi_h\| \lesssim (|\gamma| + |\gamma|^{-1})\|f\|$. In fact, it follows from the definition of the sesquilinear form $a_h(\cdot,\cdot)$ that

$$\begin{aligned}\|k\xi_h\|^2 &\lesssim -\mathrm{Re}[a_h(\xi_h,\xi_h)] + (\xi_h',\xi_h') + \|\xi_h\|_{J,k}^2\\ &\lesssim |a_h(\eta,\xi_h)| + \|\xi_h\|_{J,k}|\gamma|^{-\frac{1}{2}}(kh)^{-1}\|k\xi_h\| + \|\xi_h\|_{J,k}^2,\end{aligned}$$

where we have used an integration by parts in the second term in the right hand, i.e., $(\xi_h',\xi_h') = \sum_{j=1}^{n-1}[\xi_h']_j\overline{\xi_h}(x_j) + \xi_h'(1)\overline{\xi_h}(1)$, and the fact that $h\sum_{j=1}^n|\xi_h(x_j)|^2 \lesssim \|\xi_h\|^2$, to derive the last inequality.
From the continuity of Lemma 3.1 and (3.4) we conclude that

$$|a_h(\eta,\xi_h)| \lesssim \|\eta\|_{J,k,*}\|\xi_h\|_{J,k} \lesssim \|\eta\|_{J,k,*}^2.$$

Therefore, from (3.4) again and (3.2),

$$\begin{aligned}\|k\xi_h\|^2 &\lesssim \|\eta\|_{J,k,*}^2 + |\gamma|^{-1}(kh)^{-2}\|\eta\|_{J,k,*}^2\\ &\lesssim \big(1+|\gamma|^{-1}(kh)^{-2}\big)(|\gamma|+|\gamma|^{-1})(kh)^2\|f\|^2 \lesssim (|\gamma|+|\gamma|^{-1})^2\|f\|^2,\end{aligned}$$

which together with (3.9) proves (3.6).
By the definition of $a_h(\cdot,\cdot)$ once again and Galerkin orthogonality there holds

$$\begin{aligned}\|\xi_h'\|^2 &\lesssim \mathrm{Re}[a_h(\xi_h,\xi_h)] + \|k\xi_h\|^2 + \|\xi_h\|_{J,k}^2\\ &\lesssim |a_h(\eta,\xi_h)| + \big((|\gamma|+|\gamma|^{-1})\min(1,k^3h^2)\|f\|\big)^2\\ &\lesssim (|\gamma|+|\gamma|^{-1})(kh)^2\|f\|^2 + \big((|\gamma|+|\gamma|^{-1})\min(1,k^3h^2)\|f\|\big)^2.\end{aligned}$$

That is, (3.5) holds. This completes the proof of the theorem. ∎

Remark 3.2 (a) The L^2-error estimate can be obtained easily from (3.6) and (3.1). Both estimates exhibit the standard pollution term, but nevertheless the errors are upper bounded by data, independently of h and k. This shows that the imaginary part of the stabilization gives control of the amplitude of the wave.
(b) If the penalty term on the boundary condition is removed, i.e., if $J(u,v)$ in (2.11) is replaced by $J(u,v) := \sum_{j=1}^{n-1}\gamma h[u']_j[\overline{v}']_j$ then Theorem 3.4 still holds. This can be proved by following the analysis given in [31]. We omit the details. As we shall see in the next section, the real part of the stabilization allows us to control the phase error provided the penalty parameter is chosen appropriately.
(c) Notice that the domain Ω is of size 1. For general domain, say, $\Omega = (0,d)$, one may consider how the estimates depend on the domain size d. This can be achieved by a scaling argument. In fact, if we let $\tilde{x} = x/d$, $\tilde{k} = kd$, $\tilde{u}(\tilde{x}) = u(x)$, and $\tilde{f}(\tilde{x}) = d^2f$, then the BVP (2.1)–(2.3) becomes

$$\begin{aligned}\tilde{u}''(\tilde{x}) + \tilde{k}^2\tilde{u}(\tilde{x}) &= -\tilde{f}(\tilde{x}), \quad \tilde{x}\in(0,1)\\ \tilde{u}(0) = 0, \quad \tilde{u}'(1) - \mathbf{i}\tilde{k}\tilde{u}(1) &= 0.\end{aligned}$$

Theorem 3.4 gives the following estimate:

$$\|\tilde{u} - \tilde{u}_{\tilde{h}}\|_{1,\tilde{h}} \lesssim (|\gamma|+|\gamma|^{-1})\big(\tilde{k}\tilde{h} + \min(1,\tilde{k}^3\tilde{h}^2)\big)\|\tilde{f}\|$$

with $\tilde{h} = h/d$ and $\tilde{u}_{\tilde{h}}(\tilde{x}) = u_h(x)$. Noting that $\|\tilde{v}\|^2_{1,\tilde{h}} = d\,\|v\|^2_{1,h}$, the above estimate implies that

$$\|u - u_h\|_{1,h} \lesssim (|\gamma| + |\gamma|^{-1})\big(kdh + \min(d, k^3d^2h^2)\big)\|f\|,$$

which holds under the conditions that $kh \lesssim 1$ and $kd \gtrsim 1$.

IV. DISPERSION ANALYSIS

In this section we will consider the case where γ is a non-zero real number and $J(u,v) := \sum_{j=1}^{n-1} \gamma h[u']_j[\bar{v}']_j$ instead of the $J(u,v)$ in (2.11) (cf. Remark 3.2(b)). Using a dispersion analysis we will derive precise bounds on the error in the numerical wavenumber. These bounds are then used to prove that a particular choice of the penalty parameter allows to eliminate the pollution effect in the one-dimensional case. We remark that if $\gamma = 0$ then the CIP-FEM becomes the standard FEM which has been analyzed in [22].

A. Global FE-equations and discrete fundamental system

Let $\{\phi_1, \phi_2, \cdots, \phi_{n-1}, \phi_n\}$ be the nodal basis functions for the space V_h satisfying $\phi_j(x_l) = \delta_{jl}$, the Kronecker delta, for $j = 1, 2, \cdots, n$ and $l = 0, 1, \cdots, n$. Then the CIP-FE solution can be spanned as:

$$u_h(x) = \sum_{j=1}^{n} u_{h,j}\phi_j \quad \text{with} \quad u_{h,j} = u_h(x_j), \quad j = 1, 2, \cdots, n.$$

Let $v_h = \phi_i, i = 1, \cdots, n$ in (2.13), the CIP-FE formulation can be rewritten as the following linear system:

$$L_h U = hF, \tag{4.1}$$

where

$$L_h = h\left(a_h(\phi_j, \phi_i)\right)_{n\times n}, U = \left(u_{h,i}\right)_{n\times 1}, F = \left((f, \phi_i)\right)_{n\times 1}.$$

Denote by $t = kh$, $R = -1 - 4\gamma - t^2/6$, $S = 1 + 3\gamma - t^2/3$, we have

$$L_h = \begin{pmatrix} 2S-\gamma & R & \gamma & & & & \\ R & 2S & R & \gamma & & & \\ \gamma & R & 2S & R & \gamma & & \\ & \ddots & \ddots & \ddots & & & \\ & & \gamma & R & 2S & R & \gamma \\ & & & \gamma & R & 2S-\gamma & R+2\gamma \\ & & & & \gamma & R+2\gamma & S-2\gamma-\mathbf{i}t \end{pmatrix}. \tag{4.2}$$

Remark 4.1. The product $t = kh$ is a measure of the number of elements per wavelength (of the exact solution). In particular, if the stepwidth is such that $t = \frac{\pi}{l}$ for integer l then exactly l elements are placed on one half-wave of the exact solution.

B. Discrete wavenumber and Dispersion analysis

Recall that k is the wave number for the BVP (2.1)–(2.3) and that the functions $e^{\pm \mathbf{i}kx}$ play an important role in the solution of the BVP which satisfy the equation (2.1) with $f = 0$. The discrete wave number k_h for the CIP-FE method is defined similarly by considering the vector v with $v_j = e^{\mathbf{i}k_h jh}$ and solving the following "interior" equations:

$$\gamma v_{j-2} + Rv_{j-1} + 2Sv_j + Rv_{j+1} + \gamma v_{j+2} = 0,\ j = 3, \cdots, n-2. \tag{4.3}$$

Denote by $t_h = k_h h$, the above equations are equivalent to the equation

$$2\gamma \cos^2 t_h - \left(4\gamma + 1 + \frac{t^2}{6}\right)\cos t_h + 2\gamma + 1 - \frac{t^2}{3} = 0, \tag{4.4}$$

which has the roots

$$\cos t_h^{\pm} = \frac{4\gamma + 1 + \frac{t^2}{6} \pm \sqrt{\left(1 + \frac{t^2}{6}\right)^2 + 4\gamma t^2}}{4\gamma}. \tag{4.5}$$

For simplicity, in the following we will assume that $-1/6 \le \gamma \le 1/6$ and $t = kh \le 1$. Some simple calculations show that $|\cos t_h^-| \le 1 \le |\cos t_h^+|$. Define $k_h^- := t_h^-/h$ and $k_h^+ := t_h^+/h$. Physically, case $(-)$ describes a propagating wave whereas case $(+)$ describes a decaying wave [21].

Lemma 4.1. *Assume that $t = kh \le 1$, $-\frac{1}{6} \le \gamma \le \frac{1}{6}$, then we may show*
(i) $\left|k_h^- - k\right| \lesssim k^3h^2$;
(ii) *If $\gamma = -1/12$, then* $\left|k_h^- - k\right| \lesssim k^5h^4$;
(iii)*If $|\gamma - \gamma_o| \lesssim \frac{1}{k^2h}$ where $\gamma_o = \dfrac{6\cos t - 6 + t^2\cos t + 2t^2}{12(1-\cos t)^2}$, then* $\left|k_h^- - k\right| \lesssim kh$.

Proof. From (4.5), we have

$$1 - \cos t_h^- = \frac{t^2}{1 + \frac{t^2}{6} + \sqrt{\left(1 + \frac{t^2}{6}\right)^2 + 4\gamma t^2}} \le \frac{t^2}{2} \tag{4.6}$$

which implies $t_h^- = k_h^- h \in (0, \frac{\pi}{3}]$. Note that $\cos t_h^-$ is the function of γ and t, γ_o satisfies $\cos t_h^-(\gamma_o) = \cos t$ and hence $t_h^-(\gamma_o) = t$. From the equality in (4.6) we have,

$$\begin{aligned}
\cos t_h^- - \cos t &= \frac{t^2}{1 + \frac{t^2}{6} + \sqrt{\left(1 + \frac{t^2}{6}\right)^2 + 4\gamma_o t^2}} - \frac{t^2}{1 + \frac{t^2}{6} + \sqrt{\left(1 + \frac{t^2}{6}\right)^2 + 4\gamma t^2}} \\
&= t^2 \cdot \frac{\sqrt{\left(1 + \frac{t^2}{6}\right)^2 + 4\gamma t^2} - \sqrt{\left(1 + \frac{t^2}{6}\right)^2 + 4\gamma_o t^2}}{\left(1 + \frac{t^2}{6} + \sqrt{\left(1 + \frac{t^2}{6}\right)^2 + 4\gamma t^2}\right)\left(1 + \frac{t^2}{6} + \sqrt{\left(1 + \frac{t^2}{6}\right)^2 + 4\gamma_o t^2}\right)} \\
&= \frac{4(\gamma - \gamma_o)t^4}{\sqrt{\left(1 + \frac{t^2}{6}\right)^2 + 4\gamma t^2} + \sqrt{\left(1 + \frac{t^2}{6}\right)^2 + 4\gamma_o t^2}} \\
&\quad \times \frac{1}{\left(1 + \frac{t^2}{6} + \sqrt{\left(1 + \frac{t^2}{6}\right)^2 + 4\gamma t^2}\right)\left(1 + \frac{t^2}{6} + \sqrt{\left(1 + \frac{t^2}{6}\right)^2 + 4\gamma_o t^2}\right)}.
\end{aligned}$$

Clearly,

$$\left(1+\frac{t^2}{6}\right)^2+4\gamma t^2 \geq \left(1+\frac{t^2}{6}\right)^2+4\Big(-\frac{1}{6}\Big)t^2=\left(1-\frac{t^2}{6}\right)^2$$

On the other hand, some simple calculations show that $\left(1+\frac{t^2}{6}\right)^2+4\gamma_o t^2$ is increasing in $t\in(0,1]$ and hence

$$\left(1+\frac{t^2}{6}\right)^2+4\gamma_o t^2 \geq 1.$$

By combining the above three estimates we have

$$\left|\cos t_h^- - \cos t\right| \leq \frac{4\left|\gamma-\gamma_o\right| t^4}{\left(2-\frac{t^2}{6}\right)2\left(2+\frac{t^2}{6}\right)} \lesssim \left|\gamma-\gamma_o\right| t^4. \tag{4.7}$$

Using the inequality $\sin\theta \geq \frac{3\sqrt{3}}{2\pi}\theta$, $\forall\theta\in[0,\frac{\pi}{3}]$, we have

$$t\left|t_h^- - t\right| \leq \left|t_h^- - t\right|\left|t_h^- + t\right| \leq \frac{8\pi^2}{27}\left|2\sin\frac{t_h^- - t}{2}\sin\frac{t_h^- + t}{2}\right| = \frac{8\pi^2}{27}\left|\cos t_h^- - \cos t\right|. \tag{4.8}$$

As a consequence of (4.7) and (4.8) we have

$$\left|t_h^- - t\right| \lesssim \left|\gamma-\gamma_o\right| t^3, \tag{4.9}$$

which implies (i) and (iii) directly. It remains to prove (ii). It is easy to find that $\lim_{t\to 0}(\gamma_o+1/12)/t^2=-1/360$ which implies that $\left|(\gamma_o+1/12)/t^2\right|\lesssim 1$ for $t\in(0,1]$. Then (ii) follows from (4.9). This completes the proof of the lemma. ∎

Remark 4.2. Note that the phase difference between the exact and the linear finite element solutions is $O(k^3h^2)$ (cf. [1, 23]), while for the CIP-FEM, if the penalty parameter γ is close enough to γ_o, then the phase difference is $O(kh)$, and as a result, the CIP-FEM is pollution free (cf. Theorem 6.2 below). Figure 1 gives a plot of the optimal penalty parameter γ_o versus t for $0<t\leq 1$.

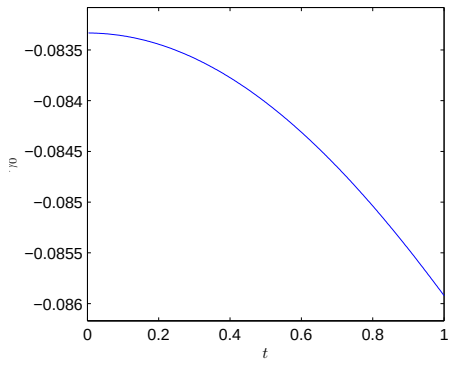

FIG. 1. The optimal penalty parameter versus $t=kh\leq 1$.

V. THE DISCRETE GREEN'S FUNCTION

To construct the discrete Green's function, we first find the inverse of the stiffness matrix L_h. Inspired by the formulation of the Green's function for the BVP (cf. (2.4)), we find $G_h = L_h^{-1}$ of the following form:

$$G_{h,j,m} = \begin{cases} \sum_{i=1}^4 A_{m,i}\eta_i^j, & j < m, \\ \sum_{i=1}^4 B_{m,i}\eta_i^j, & j \geq m, \end{cases} \tag{5.1}$$

where $\eta_1 = e^{-\mathbf{i}k_h^- h}, \eta_2 = e^{\mathbf{i}k_h^- h}, \eta_3 = e^{-\mathbf{i}k_h^+ h}, \eta_4 = e^{\mathbf{i}k_h^+ h}$.
By the definition of $\eta_i, i = 1, 2, 3, 4$ there hold the facts:

$$\eta_1\eta_2 = \eta_3\eta_4 = 1, \qquad \eta_1 + \eta_2 = 2\cos t_h^-, \qquad \eta_3 + \eta_4 = 2\cos t_h^+. \tag{5.2}$$

If $|\gamma| \leq 1/6$, from (4.5) and some simple calculations, we can get

$$|\cos t_h^+ - 1| \geq 3. \tag{5.3}$$

Without loss of generality, assume $|\eta_4| > |\eta_3|$, it is clear that

$$|\eta_4| > 3 \ \text{ and } \ |\eta_3| < \frac{1}{3}. \tag{5.4}$$

From (5.1), the solution of (4.1) is represented as

$$u_{h,j} = h\sum_{m=1}^{n} G_{h,j,m}(f, \phi_m), \quad j = 1, 2, \cdots, n, \tag{5.5}$$

and hence the CIP-FE solution is given by

$$u_h = \sum_{j=1}^{n} u_{h,j}\phi_j.$$

To represent the derivative of the CIP-FE solution, we define a $n \times n$ matrix H_h as

$$H_{h,j,m} = G_{h,j,m} - G_{h,j-1,m} \ \ 1 \leq j \leq n, \ \text{ Here } \ G_{h,0,m} := 0. \tag{5.6}$$

It is clear that

$$u_h'(x) = \frac{u_{h,j} - u_{h,j-1}}{h} = \sum_{m=1}^{n} H_{h,j,m}(f, \phi_m), \quad \forall x \in (x_{j-1}, x_j), \ j = 1, \cdots, n. \tag{5.7}$$

Throughout this section let $\widetilde{C}$ denote a *general function* that may have different expressions at different places but is bounded (uniformly) by some constant independent of k, h, and the penalty parameters. Note that $\widetilde{C}$ is allowed to be complex-valued. We first state a simple but useful lemma without proof.

Lemma 5.1. *Suppose $0 < t \leq 1$, if $|b| \leq \sigma_1|a|$, $0 < \sigma_1 < 1$, a, b and σ_1 are independent of the penalty parameter. Then*

$$\frac{1}{a - bt} = \frac{1}{a}\left(1 + \widetilde{C}t\right). \tag{5.8}$$

The following lemma presents estimates for $H_{h,j,m}$.

Lemma 5.2. *Assume that $t = kh \le 1$, $k \ge 1$, $0 < |\gamma| \le \frac{1}{6}$. Then*

$$H_{h,j,m} = \begin{cases} \cos(jt_h^-)e^{\mathbf{i}mt_h^-} + \widetilde{C}t + \widetilde{C}\eta_4^{j-m}, & j < m, \\ \mathbf{i}\sin(mt_h^-)e^{\mathbf{i}jt_h^-} + \widetilde{C}t + \widetilde{C}\eta_4^{m-j}, & j \ge m, \end{cases} \tag{5.9}$$

where $\widetilde{C}$ is a general function which is bounded by some constant independent of k, h, and the penalty parameters.

Proof. The proof is divided into four steps.

Step 1. Solving for $A_{m,i}$ and $B_{m,i}$. $G_{h,j,m}$ are determined by the system of equations:

$$\begin{cases} (2S-\gamma)G_{h,1,m} + RG_{h,2,m} + \gamma G_{h,3,m} = \delta_{1,m}, \\ RG_{h,1,m} + 2SG_{h,2,m} + RG_{h,3,m} + \gamma G_{h,4,m} = \delta_{2,m}, \\ \gamma G_{h,n-3,m} + RG_{h,n-2,m} + (2S-\gamma)G_{h,n-1,m} + (R+2\gamma)G_{h,n,m} = \delta_{n-1,m}, \\ \gamma G_{h,n-2,m} + (R+2\gamma)G_{h,n-1,m} + (S-2\gamma-\mathbf{i}t)G_{h,n,m} = \delta_{n,m}, \\ \gamma G_{h,j-2,m} + RG_{h,j-1,m} + 2SG_{h,j,m} + RG_{h,j+1,m} + \gamma G_{h,j+2,m} = \delta_{j,m}, \end{cases} \tag{5.10}$$

where $3 \le j \le n-2$ in the last equality of the above system and $\delta_{j,m}$, $1 \le j, m \le n$, is the Kronecker delta.

Formula (4.3) yields

$$\gamma\eta_i^{-2} + R\eta_i^{-1} + 2S + R\eta_i + \gamma\eta_i^2 = 0. \tag{5.11}$$

We first consider $m = 5, \cdots, n-3$. From (5.1), the system (5.10) is reduced to the following system of eight equations:

$$\begin{cases} \sum_{i=1}^4 \eta_i(2S - \gamma + R\eta_i + \gamma\eta_i^2)A_{m,i} = 0, \\ \sum_{i=1}^4 \eta_i^2(R\eta_i^{-1} + 2S + R\eta_i + \gamma\eta_i^2)A_{m,i} = 0, \\ \sum_{i=1}^4 \eta_i^{m-2}\left[(\gamma\eta_i^{-2} + R\eta_i^{-1} + 2S + R\eta_i)A_{m,i} + (\gamma\eta_i^2)B_{m,i}\right] = 0, \\ \sum_{i=1}^4 \eta_i^{m-1}\left[(\gamma\eta_i^{-2} + R\eta_i^{-1} + 2S)A_{m,i} + (R\eta_i + \gamma\eta_i^2)B_{m,i}\right] = 0, \\ \sum_{i=1}^4 \eta_i^{m}\left[(\gamma\eta_i^{-2} + R\eta_i^{-1})A_{m,i} + (2S + R\eta_i + \gamma\eta_i^2)B_{m,i}\right] = 1, \\ \sum_{i=1}^4 \eta_i^{m+1}\left[(\gamma\eta_i^{-2})A_{m,i} + (R\eta_i^{-1} + 2S + R\eta_i + \gamma\eta_i^2)B_{m,i}\right] = 0, \\ \sum_{i=1}^4 \eta_i^{n-1}\left[\gamma\eta_i^{-2} + R\eta_i^{-1} + 2S - \gamma + (R+2\gamma)\eta_i\right]B_{m,i} = 0, \\ \sum_{i=1}^4 \eta_i^{n}\left[\gamma\eta_i^{-2} + (R+2\gamma)\eta_i^{-1} + S - 2\gamma - \mathbf{i}t\right]B_{m,i} = 0. \end{cases} \tag{5.12}$$

Plugging (5.11) into the first seven equations of (5.12) gives

$$\begin{aligned} &\textstyle\sum_{i=1}^4 (\gamma\eta_i^{-1} + R + \gamma\eta_i)A_{m,i} = 0, \\ &\textstyle\sum_{i=1}^4 \gamma A_{m,i} = 0, \\ &\textstyle\sum_{i=1}^4 \gamma\eta_i^m(B_{m,i} - A_{m,i}) = 0, \\ &\textstyle\sum_{i=1}^4 (R + \gamma\eta_i)\eta_i^m(B_{m,i} - A_{m,i}) = 0, \\ &\textstyle\sum_{i=1}^4 (\gamma\eta_i^{-2} + R\eta_i^{-1})\eta_i^m(A_{m,i} - B_{m,i}) = 1, \\ &\textstyle\sum_{i=1}^4 \gamma\eta_i^{m-1}(A_{m,i} - B_{m,i}) = 0, \\ &\textstyle\sum_{i=1}^4 \gamma(\eta_i^{-1} - 2 + \eta_i)\eta_i^n B_{m,i} = 0. \end{aligned} \tag{5.13}$$

By $R = -1 - 4\gamma - t^2/6$, $S = 1 + 3\gamma - t^2/3$, the eighth equation of (5.12) yields

$$\sum_{i=1}^4 \Big(1 - \frac{t^2}{3} - \Big(1 + \frac{t^2}{6}\Big)\eta_i^{-1} + \gamma(1-\eta_i^{-1})^2 - \mathbf{i}t\Big)\eta_i^n B_{m,i} = 0. \tag{5.14}$$

Then, by simplifying (5.13) and (5.14), a 8×8 system which is equivalent to the system (5.12) can be obtained:

$$\begin{bmatrix} -U_m & U_m \\ V_1 & V_2 \end{bmatrix} \begin{bmatrix} A_m \\ B_m \end{bmatrix} = \begin{bmatrix} z \\ 0 \end{bmatrix} \qquad z = [-1/\gamma, 0, 0, 0]^T, \tag{5.15}$$

where $A_m = [A_{m,1}, A_{m,2}, A_{m,3}, A_{m,4}]^T$, $B_m = [B_{m,1}, B_{m,2}, B_{m,3}, B_{m,4}]^T$, and the i-th $(i = 1, 2, 3, 4)$ column of the matrix U_m, V_1, V_2 are stated as follows:

$$U_m(:, i) = \eta_i^m \begin{pmatrix} \eta_i^{-2} \\ \eta_i^{-1} \\ 1 \\ \eta_i \end{pmatrix}, \qquad V_1(:, i) = \begin{pmatrix} \eta_i^{-1} + \eta_i \\ 1 \\ 0 \\ 0 \end{pmatrix}, \qquad V_2(:, i) = \begin{pmatrix} 0 \\ 0 \\ a_i \\ b_i \end{pmatrix},$$

where

$$a_i = (\eta_i^{-1} - 2 + \eta_i)\eta_i^n, \tag{5.16}$$

$$b_i = \Big(1 - \frac{t^2}{3} - \Big(1 + \frac{t^2}{6}\Big)\eta_i^{-1} + \gamma(1 - \eta_i^{-1})^2 - \mathbf{i}t\Big)\eta_i^n, \qquad i = 1, 2, 3, 4. \tag{5.17}$$

Next we consider $m = 2, 3, 4, n-2, n-1, n$. Then the linear system is underdetermined since it contains less than 8 equations, however, we can show that the system (5.15) gives a special solution. We only prove the case $m = 2$, other cases $(m = 3, 4, n-2, n-1, n)$ can be obtained similarly, we leave the derivation to the interested reader. When $m = 2$, from (5.1) and (5.11), the system (5.10) is reduced to the following system of five equations:

$$\begin{cases} \sum_{i=1}^4 \eta_i \left[(2S - \gamma)A_{2,i} + (R\eta_i + \gamma\eta_i^2)B_{2,i}\right] = 0, \\ \sum_{i=1}^4 \eta_i^2 \left[R\eta_i^{-1}A_{2,i} + (2S + R\eta_i + \gamma\eta_i^2)B_{2,i}\right] = 1, \\ \sum_{i=1}^4 \eta_i^3 \left[(\gamma\eta_i^{-2})A_{2,i} + (R\eta_i^{-1} + 2S + R\eta_i + \gamma\eta_i^2)B_{2,i}\right] = 0, \\ \sum_{i=1}^4 \eta_i^{n-1} \left[\gamma\eta_i^{-2} + R\eta_i^{-1} + 2S - \gamma + (R + 2\gamma)\eta_i\right] B_{2,i} = 0, \\ \sum_{i=1}^4 \eta_i^n \left[\gamma\eta_i^{-2} + (R + 2\gamma)\eta_i^{-1} + S - 2\gamma - \mathbf{i}t\right] B_{2,i} = 0. \end{cases} \tag{5.18}$$

We remark that, although the above system is underdetermined, $G_{h,j,2}$ is uniquely determined by (5.1). As a matter of fact, (5.18) can be viewed as a system of five unknowns $B_{2,i}, i = 1, 2, 3, 4$ and $\sum_{i=1}^4 \eta_i A_{2,i}$. As we just mentioned, a solution of (5.18) can be obtained from (5.15) with $m = 2$, because of the following facts. The last three equations of (5.18) are the same as the last three equations of (5.12) (with $m = 2$). The first equation of (5.18) can be obtained from the sum of the first equation of (5.12) and the fourth equation of (5.13) (with $m = 2$). Similarly, the second equation of (5.18) can be obtained by substracting the second equation of (5.13) from the fifth equation of (5.12) (with $m = 2$).

For $m = 1$, the system (5.10) is reduced to the system of four equations:
$(V_1 + V_2)B_1 = z$.
Denote by $V = V_1 + V_2$. We will show in Step 3 of this proof that V is invertible. As a consequence, $B_1 = V^{-1}z$. Noting that U_m is also invertible, we solve (5.15) to obtain $A_m = -V^{-1}V_2U_m^{-1}z$, $B_m = V^{-1}V_1U_m^{-1}z$, $1 < m \leq n$.

Step 2. Estimating a_i and b_i. In order to estimate A_m and B_m, we prove the following assertions in this step:

$$|a_1| = |a_2| \le t^2, \quad a_3 = a_4\eta_4^{-2n}, \quad |a_4| \ge 6|\eta_4|^n, \tag{5.19}$$

$$|b_1| > \frac{5}{3}t, \quad |b_2| < \frac{t^2}{3}, \quad |b_3| < \frac{2}{3}|\eta_4|^{1-n}, \quad |b_4| < \frac{3}{2}|\eta_4|^n, \tag{5.20}$$

$$|a_1b_2 - a_2b_1| = \left|t^2(\eta_1 - \eta_2)\right| \le 2t^2, \quad |a_3b_4 - b_3a_4| \le 2t^2\,|\eta_4|^{-n}\,|a_4|\,. \tag{5.21}$$

where η_4 satisfies (5.4).
It follows from (4.6) that

$$|a_1| = \left|2\cos t_h^- - 2\right||\eta_1^n| \le t^2, \quad |a_2| = \left|2\cos t_h^- - 2\right||\eta_2^n| \le t^2.$$

Using the identity $\eta_3 = \eta_4^{-1}$ and (5.16) we get

$$a_3 = \eta_3^n(\eta_3 + \eta_4 - 2) = \eta_4^{-2n}a_4.$$

It follows from (5.3) and (5.16) that

$$|a_4| = |\eta_4^n(\eta_3 + \eta_4 - 2)| = |\eta_4^n|\left|2\cos t_h^+ - 2\right| \ge 6\,|\eta_4^n|\,.$$

Therefore (5.19) holds.
Next, we turn to prove (5.20). Noting that $0 < |\gamma| < 1/6$, from (5.17), (5.2), and (4.5) we have

$$\begin{aligned}
|b_2| &= \left|1 - \frac{t^2}{3} - \Big(1 + \frac{t^2}{6}\Big)\eta_2^{-1} + \gamma(2\cos t_h^- - 2)\eta_2^{-1} - \mathbf{i}t\right| \\
&= \left|1 - \frac{t^2}{3} - \Big(1 + \frac{t^2}{6}\Big)\eta_2^{-1} + \frac{1 + \frac{t^2}{6} - \sqrt{\big(1 + \frac{t^2}{6}\big)^2 + 4\gamma t^2}}{2}\eta_2^{-1} - \mathbf{i}t\right| \\
&\le \left|1 - \frac{t^2}{3} - \frac{1 + \frac{t^2}{6} + \sqrt{\big(1 + \frac{t^2}{6}\big)^2 + 4\gamma t^2}}{2}\cos t_h^-\right| \\
&\qquad + \left|\frac{1 + \frac{t^2}{6} + \sqrt{\big(1 + \frac{t^2}{6}\big)^2 + 4\gamma t^2}}{2}\sin t_h^- - t\right| := \text{(I)} + \text{(II)},
\end{aligned}$$

where

$$\begin{aligned}
\text{(I)} &= \left|1 - \frac{t^2}{3} - \frac{1 + \frac{t^2}{6} + \sqrt{\big(1 + \frac{t^2}{6}\big)^2 + 4\gamma t^2}}{2}\Big(1 - \frac{t^2}{1 + \frac{t^2}{6} + \sqrt{\big(1 + \frac{t^2}{6}\big)^2 + 4\gamma t^2}}\Big)\right| \\
&= \left|\frac{\sqrt{\big(1 + \frac{t^2}{6}\big)^2 + 4\gamma t^2} - 1 - \frac{t^2}{6}}{2}\right| \le \frac{t^2}{6},
\end{aligned}$$

$$
\begin{aligned}
\text{(II)} &= \left| \frac{1+\frac{t^2}{6}+\sqrt{\left(1+\frac{t^2}{6}\right)^2+4\gamma t^2}}{2}\sqrt{1-(\cos t_h^-)^2}-t \right| \\
&= \left| \frac{\sqrt{2t^2\left(1-\frac{t^2}{3}+\sqrt{\left(1+\frac{t^2}{6}\right)^2+4\gamma t^2}\right)}-2t}{2} \right| \\
&= \frac{\left|-\frac{1}{3}-\frac{t^2}{12}+4\gamma\right| t^3}{\left(\sqrt{2\left(1-\frac{t^2}{3}+\sqrt{\left(1+\frac{t^2}{6}\right)^2+4\gamma t^2}\right)}+2\right)\left(1+\frac{t^2}{3}+\sqrt{\left(1+\frac{t^2}{6}\right)^2+4\gamma t^2}\right)} < \frac{t^3}{6},
\end{aligned}
$$

we therefore arrive at

$$
|b_2| \le \text{(I)}+\text{(II)} < \frac{t^2}{3}. \tag{5.22}
$$

Noting that $\bar{\eta}_2 = \eta_1$, it is clear that

$$
b_1 = \eta_1^n\Big(1-\frac{t^2}{3}-\Big(1+\frac{t^2}{6}\Big)\eta_1^{-1}+\gamma(1-\eta_1^{-1})^2-\mathbf{i}t\Big) = \bar{b}_2 - 2\mathbf{i}t\eta_1^n.
$$

Obviously, $|b_1| \ge 2t - |b_2| > \frac{5}{3}t$. From (5.4),

$$
\begin{aligned}
|b_3| &= \left|\eta_3^n\Big(1-\frac{t^2}{3}-\Big(1+\frac{t^2}{6}\Big)\eta_3^{-1}+\gamma(1-\eta_3^{-1})^2-\mathbf{i}t\Big)\right| \\
&= \left|\eta_3^n\Big(1-\frac{t^2}{3}-\Big(1+\frac{t^2}{6}\Big)\eta_3^{-1}+\gamma(2\cos t_h^+-2)\eta_3^{-1}-\mathbf{i}t\Big)\right| \\
&= |\eta_3^n|\left|1-\frac{t^2}{3}+\frac{2\gamma t^2}{1+\frac{t^2}{6}+\sqrt{\left(1+\frac{t^2}{6}\right)^2+4\gamma t^2}}\eta_3^{-1}-\mathbf{i}t\right| \\
&\le |\eta_3^{n-1}|\left(\frac{1}{3}\left|1-\frac{t^2}{3}-\mathbf{i}t\right|+\frac{t^2}{6}\right) < \frac{2}{3}|\eta_3|^{n-1} = \frac{2}{3}|\eta_4|^{1-n}.
\end{aligned}
$$

Similarly,

$$
|b_4| = |\eta_4^n|\left|1-\frac{t^2}{3}+\frac{2\gamma t^2}{1+\frac{t^2}{6}+\sqrt{\left(1+\frac{t^2}{6}\right)^2+4\gamma t^2}}\eta_4^{-1}-\mathbf{i}t\right| < \frac{3}{2}|\eta_4^n|.
$$

This completes the proof of (5.20).
It remains to prove (5.21). We derive from (5.16)–(5.17), (5.2), and (4.5) that

$$
\begin{aligned}
|a_1b_2-a_2b_1| &= \Big|(\eta_1+\eta_2-2)\Big(1-\frac{t^2}{3}-\Big(1+\frac{t^2}{6}\Big)\eta_1+\gamma(1-\eta_1)^2-\mathbf{i}t\Big) \\
&\quad -(\eta_1+\eta_2-2)\Big(1-\frac{t^2}{3}-\Big(1+\frac{t^2}{6}\Big)\eta_2+\gamma(1-\eta_2)^2-\mathbf{i}t\Big)\Big| \\
&= \Big|(\eta_1+\eta_2-2)\Big(\gamma(\eta_1+\eta_2-2)(\eta_1-\eta_2)-\Big(1+\frac{t^2}{6}\Big)(\eta_1-\eta_2)\Big)\Big| \\
&= \big|t^2(\eta_1-\eta_2)\big| \le 2t^2.
\end{aligned}
$$

Similarly,

$$a_3b_4 - b_3a_4 = t^2(\eta_3 - \eta_4) = t^2 a_4\eta_4^{-n}\frac{\eta_3 - \eta_4}{\eta_3 + \eta_4 - 2} = t^2 a_4\eta_4^{-n}\frac{1 - \eta_4^2}{1 + \eta_4^2 - 2\eta_4},$$

hence, again from (5.4),

$$|a_3b_4 - b_3a_4| = t^2\left|a_4\eta_4^{-n}\frac{1 - \eta_4}{1 + \eta_4}\right| \le t^2\left|a_4\eta_4^{-n}\right|\frac{|\eta_4| + 1}{|\eta_4| - 1} \le 2t^2\left|\eta_4\right|^{-n}|a_4|.$$

This completes the proof of (5.21).

Step 3. Estimating A_m and B_m. Since $t = kh \le 1$ and $k \ge 1$, from (5.4), we have

$$|\eta_4|^{-n} < \left(\frac{1}{3}\right)^{\frac{1}{h}} \le \left(\frac{1}{3}\right)^{\frac{1}{t}} \le \frac{1}{3}t. \tag{5.23}$$

Next we estimate $\dfrac{1}{\det V}$, where $V = V_1 + V_2$. By some simple calculation, we have

$$\det V = [(\eta_3 + \eta_4) - (\eta_1 + \eta_2)][(a_2 - a_1)(b_4 - b_3) - (b_2 - b_1)(a_4 - a_3)] \tag{5.24}$$

where a_i and b_i are defined by (5.16) and (5.17), respectively. We analyze and estimate each term of $\det V$. From (5.20), it is clear that $|\dfrac{b_2}{b_1}| < \dfrac{t}{5}$. Hence,

$$b_1 - b_2 = b_1(1 + \theta_1 t) \tag{5.25}$$

where θ_1 is a general function satisfying $|\theta_1| < \frac{1}{5}$.
It follows from (5.23), (5.19), and (5.25) that

$$(b_1 - b_2)(a_4 - a_3) = b_1a_4(1 + \theta_2 t) \tag{5.26}$$

where θ_2 is a general function and $|\theta_2| < \frac{1}{3}$.
From (5.19)–(5.20) and (5.23), we have

$$|(a_2 - a_1)(b_4 - b_3)| \le \frac{2}{3}t^2|a_4| \le \frac{2}{5}t|b_1a_4|.$$

It follows from (5.24), (5.26), and the above inequality that

$$\det V = b_1a_4[(\eta_3 + \eta_4) - (\eta_1 + \eta_2)](1 + \theta_3 t),$$

where θ_3 is a general function and $|\theta_3| < \dfrac{11}{15}$. Therefore from Lemma 5.1,

$$\frac{1}{\det V} = \frac{1 + \widetilde{C}t}{b_1a_4\sigma}, \tag{5.27}$$

where $\sigma := \eta_3 + \eta_4 - (\eta_1 + \eta_2)$. Note from (4.5) and (5.2) that

$$\frac{1}{\sigma} = \frac{\gamma}{\sqrt{\left(1 + \frac{t^2}{6}\right)^2 + 4\gamma t^2}} = \gamma(1 + \widetilde{C}t^2). \tag{5.28}$$

In order to estimate B_1, we consider the first column of V^*, the adjugate of V. From (5.23) and (5.19)–(5.21), by some calculations, we have

$$V^*(:,1) = \begin{pmatrix} a_3b_4 - b_3a_4 + b_2a_4 - a_2b_4 + a_2b_3 - b_2a_3 \\ -b_1a_4 + a_1b_4 + b_3a_4 - a_3b_4 + b_1a_3 - a_1b_3 \\ b_1a_4 - a_1b_4 + a_2b_4 - b_2a_4 + a_1b_2 - a_2b_1 \\ a_3b_2 - a_2b_3 + a_1b_3 - b_1a_3 + a_2b_1 - a_1b_2 \end{pmatrix} = \begin{pmatrix} b_1a_4\widetilde{C}t \\ -b_1a_4(1 + \widetilde{C}t) \\ b_1a_4(1 + \widetilde{C}t) \\ \eta_4^{-n}b_1a_4\widetilde{C}t \end{pmatrix},$$

hence, from (5.27) and (5.28),

$$B_1 = V^{-1}z = \frac{1}{\det V}V^*z = \frac{1}{\det V}\Big(-\frac{1}{\gamma}\Big)V^*(:,1) = \begin{pmatrix} \widetilde{C}t \\ 1+\widetilde{C}t \\ -1+\widetilde{C}t \\ \eta_4^{-n}\widetilde{C}t \end{pmatrix}. \tag{5.29}$$

We turn to estimate A_m and B_m for $m > 1$. It follows from the definitions of U_m and z that,

$$U_m^{-1}z = -\frac{1}{\gamma}\begin{pmatrix} \dfrac{\eta_1^{2-m}\eta_2\eta_3\eta_4}{(\eta_3-\eta_1)(\eta_4-\eta_1)(\eta_2-\eta_1)} \\ \dfrac{\eta_2^{2-m}\eta_1\eta_3\eta_4}{(\eta_3-\eta_2)(\eta_4-\eta_2)(\eta_1-\eta_2)} \\ \dfrac{\eta_3^{2-m}\eta_1\eta_2\eta_4}{(\eta_1-\eta_3)(\eta_2-\eta_3)(\eta_4-\eta_3)} \\ \dfrac{\eta_4^{2-m}\eta_1\eta_2\eta_3}{(\eta_1-\eta_4)(\eta_2-\eta_4)(\eta_3-\eta_4)} \end{pmatrix} = (1+\widetilde{C}t^2)\begin{pmatrix} \dfrac{\eta_2^m}{\eta_2-\eta_1} \\ \dfrac{\eta_1^m}{\eta_1-\eta_2} \\ \dfrac{\eta_4^m}{\eta_3-\eta_4} \\ \dfrac{\eta_3^m}{\eta_4-\eta_3} \end{pmatrix}, \tag{5.30}$$

where we have used (5.2) and (5.28) to derive the last equality.
Next we estimate V^*V_1. Clearly, $V_1(:,2) = V_1(:,1)$, $V_1(:,4) = V_1(:,3)$, and so is V^*V_1. It follows from (5.19)–(5.21) and (5.23) that,

$$\begin{aligned} V^*V_1(:,[1,3]) &= V^*V_1(:,[2,4]) \\ &= \sigma\begin{pmatrix} a_2b_4 - a_2b_3 - b_2a_4 + b_2a_3 & a_3b_4 - b_3a_4 \\ a_1b_3 - a_1b_4 + b_1a_4 - b_1a_3 & b_3a_4 - a_3b_4 \\ a_2b_1 - a_1b_2 & a_2b_4 - a_1b_4 - a_4b_2 + a_4b_1 \\ a_1b_2 - b_1a_2 & a_3b_2 - a_3b_1 - b_3a_2 + b_3a_1 \end{pmatrix} \\ &= \sigma b_1 a_4\begin{pmatrix} \widetilde{C}t & \eta_4^{-n}\widetilde{C}t \\ 1+\widetilde{C}t & \eta_4^{-n}\widetilde{C}t \\ \eta_4^{-n}(\eta_1-\eta_2)\widetilde{C}t & 1+\widetilde{C}t \\ \eta_4^{-n}(\eta_1-\eta_2)\widetilde{C}t & -\eta_4^{-2n}(1-\eta_4\widetilde{C}t) \end{pmatrix}. \end{aligned} \tag{5.31}$$

From (5.27), (5.30), (5.31), (5.4), and $|\eta_1| = |\eta_2| = 1$, we have

$$\begin{aligned} B_m = V^{-1}V_1U_m^{-1}z &= \frac{1}{\det V}V^*V_1U_m^{-1}z \\ &= (1+\widetilde{C}t)\begin{pmatrix} \dfrac{\eta_2^m-\eta_1^m}{\eta_2-\eta_1}\widetilde{C}t + \widetilde{C}t \\ \dfrac{\eta_2^m-\eta_1^m}{\eta_2-\eta_1}(1+\widetilde{C}t) + \widetilde{C}t \\ \eta_4^{-n}\widetilde{C}t + \eta_4^{m-1}\widetilde{C} \\ \eta_4^{-n}\widetilde{C}t + \eta_4^{-2n-1+m}\widetilde{C} \end{pmatrix} = \begin{pmatrix} \dfrac{\widetilde{C}t}{\sin t_h^-} \\ \dfrac{\sin(mt_h^-)+\widetilde{C}t}{\sin t_h^-} \\ \eta_4^{m-1}\widetilde{C} \\ \eta_4^{-n}\widetilde{C}t + \eta_4^{-2n-1+m}\widetilde{C} \end{pmatrix}. \end{aligned} \tag{5.32}$$

By some calculations, we find that

$$V^*(1,i) = -V^*(2,i),\ V^*(3,i) = -V^*(4,i),\ \text{where } i = 3,4.$$

Since $V_2(1,:) = V_2(2,:) = 0$, we have $(V^*V_2)([1,3],:) = -(V^*V_2)([2,4],:)$. Therefore, similar to (5.31), from (5.23), (5.27) and (5.19)–(5.21), we may show that

$$\begin{aligned}&(V^*V_2)([1,3],:) = -(V^*V_2)([2,4],:) \qquad (5.33)\\ =&\sigma b_1 a_4 \begin{pmatrix} 1+\widetilde{C}t & \widetilde{C}t & \eta_4^{-n}\widetilde{C}t & \eta_4^{-n}\widetilde{C}t \\ \eta_4^{-n}(\eta_1-\eta_2)\widetilde{C}t & \eta_4^{-n}(\eta_1-\eta_2)\widetilde{C}t & \eta_4^{-2n}(\eta_4\widetilde{C}t-1) & -1+\widetilde{C}t \end{pmatrix}.\end{aligned}$$

It follows from (5.27), (5.30) and (5.33) that,

$$\begin{aligned}A_m([1,3]) = -A_m([2,4]) &= -\frac{1}{\det V}(V^*V_2)([1,3],:)U_m^{-1}z \qquad (5.34)\\ &= \begin{pmatrix} \dfrac{\eta_2^m}{\eta_1-\eta_2} + \dfrac{\widetilde{C}t}{\sin t_h^-} \\ \eta_4^{-m}\widetilde{C}t + \eta_4^{-m-1}\widetilde{C} \end{pmatrix}.\end{aligned}$$

Step 4. Finishing up. It is time to consider $H_{h,j,m}$. Let $w_1^T = [\eta_1, \eta_2, \eta_3, \eta_4]$, $w_j^T = [(\eta_1-1)\eta_1^{j-1}, (\eta_2-1)\eta_2^{j-1}, (\eta_3-1)\eta_3^{j-1}, (\eta_4-1)\eta_4^{j-1}]$ for $j>1$. From (5.1), (5.6), (5.32), and (5.34), we have, for $m=1$,

$$\begin{aligned}H_{h,1,1} &= G_{h,1,1} = w_1^T B_1 = e^{\mathbf{i}t_h^-} + \widetilde{C}\eta_4^{-1} + \widetilde{C}t = \widetilde{C} = \mathbf{i}\sin(t_h^-)e^{\mathbf{i}t_h^-} + \widetilde{C}t + \widetilde{C},\\ H_{h,j,1} &= G_{h,j,1} - G_{h,j-1,1} = w_j^T B_1 = \mathbf{i}\sin(t_h^-)e^{\mathbf{i}jt_h^-} + \widetilde{C}t + \widetilde{C}\eta_4^{1-j}, \quad j>1,\end{aligned}$$

and for $m>1$,

$$\begin{aligned}H_{h,1,m} &= G_{h,1,m} = w_1^T A_m = \cos(t_h^-)e^{\mathbf{i}mt_h^-} + \widetilde{C}t + \widetilde{C}\eta_4^{1-m},\\ H_{h,j,m} &= G_{h,j,m} - G_{h,j-1,m} = w_j^T A_m = \cos(jt_h^-)e^{\mathbf{i}mt_h^-} + \widetilde{C}t + \widetilde{C}\eta_4^{j-m}, \quad 1<j<m,\\ H_{h,j,m} &= G_{h,j,m} - G_{h,j-1,m} = w_j^T B_m = \mathbf{i}\sin(mt_h^-)e^{\mathbf{i}jt_h^-} + \widetilde{C}t + \widetilde{C}\eta_4^{m-j}, \quad j>m,\\ H_{h,m,m} &= G_{h,m,m} - G_{h,m-1,m} = \sum_{i=1}^4 B_{m,i}\eta_i^m - \sum_{i=1}^4 A_{m,i}\eta_i^{m-1}\\ &= w_m^T B_m + \sum_{i=1}^4 (B_{m,i}-A_{m,i})\eta_i^{m-1} = w_m^T B_m\\ &= \mathbf{i}\sin(mt_h^-)e^{\mathbf{i}mt_h^-} + \widetilde{C}t + \widetilde{C},\end{aligned}$$

where we have used $\sum_{i=1}^4 (B_{m,i}-A_{m,i})\eta_i^{m-1} = 0$ (cf. the sixth equation in (5.13)). This completes the proof of the lemma. ∎

From Lemma 5.2 and (5.7), we have

$$\begin{aligned}u_h'(x) = \sum_{m=1}^n H_{h,j,m}(f,\phi_m) &= \sum_{m=1}^j \mathbf{i}\sin(mt_h^-)e^{\mathbf{i}jt_h^-}(f,\phi_m) \qquad (5.35)\\ &+ \sum_{m=j+1}^n \cos(jt_h^-)e^{\mathbf{i}mt_h^-}(f,\phi_m) + t\sum_{m=1}^n \widetilde{C}(f,\phi_m)\\ &+ \sum_{m=1}^n \eta_4^{-|m-j|}\widetilde{C}(f,\phi_m), \quad \forall x\in(x_{j-1},x_j),\ 1\le j\le n.\end{aligned}$$

Comparing with the continuous case (2.6) we see that the first two contributions in the right hand side of (5.35) consist of the discrete travelling wave, whereas the last two perturbed terms will be shown to be of the same order as the interpolation error.

VI. STABILITY AND PREASYMPTOTIC ERROR ESTIMATES FOR THE CIP-FEM

In this section, we consider the stability and error estimates of the CIP-FE solution in the discrete semi-norm $\|\cdot\|_{1,h}$ (defined in (2.14)) for real penalty parameters.

Theorem 6.1. *Under the conditions of Lemma 5.2, the CIP-FEM* (2.13) *attains a unique solution u_h that satisfies the stability estimate*

$$\|u_h\|_{1,h} \lesssim \|f\|. \tag{6.1}$$

Proof. Let us estimate each term in the definition of $\|\cdot\|_{1,h}$(cf. (2.14)). First, from (5.35), it is clear that

$$|u_h'(x)| \lesssim \sum_{m=1}^{n} |(f,\phi_m)| \lesssim \|f\|, \quad \forall x \in (x_{j-1},x_j), \quad j=1,\cdots,n,$$

and hence,

$$\|u_h'\| \lesssim \|f\|. \tag{6.2}$$

Secondly,

$$|[u_h']_j| = |u_h'(x_j+) - u_h'(x_j-)| \le |u_h'(x_j+)| + |u_h'(x_j-)| \lesssim \|f\|,$$

which implies

$$\sum_{j=1}^{n-1} |\gamma| h \left|[u_h']_j\right|^2 \lesssim \|f\|^2.$$

Therefore,

$$\|u_h\|_{1,h} = \left(|u_h|_{1,h}^2 + \sum_{j=1}^{n-1} |\gamma| h \left|[u_h']_j\right|^2 \right)^{\frac{1}{2}} \lesssim \|f\|.$$

This completes the proof of Theorem 6.1. ∎

Remark 6.1. This stability estimate for the CIP-FEM (as well as FEM) is of the same order as that of the continuous problem (cf. (2.8)). Note that the estimate holds for real penalty parameters in $[-\frac{1}{6},\frac{1}{6}]$ under the condition $kh \le 1$ in current one-dimensional setting. The same result has been proved for the one-dimensional FEM in [22]. For stability estimates of the CIP-FEM for higher-dimensional problems, we refer to [31] which, particularly, gives estimates for imaginary penalty parameters under the condition $kh \lesssim 1$.

Theorem 6.2. *Under the conditions of Lemma 5.2,*

$$\|u - u_h\|_{1,h} \lesssim (kh + \left|k_h^- - k\right|) \|f\| \lesssim (kh + k^3h^2) \|f\|. \tag{6.3}$$

Furthermore, if $\gamma = -\frac{1}{12}$, *then*

$$\|u - u_h\|_{1,h} \lesssim (kh + k^5h^4)\,\|f\|\,. \tag{6.4}$$

If $|\gamma - \gamma_o| \lesssim \frac{1}{k^2h}$, *then*

$$\|u - u_h\|_{1,h} \lesssim kh\,\|f\|\,. \tag{6.5}$$

Here γ_o *is defined in Lemma 4.1.*

Proof. Suppose $n \lesssim k^2$, that is, $k^2h \gtrsim 1$, otherwise, (6.5) is proved by using the Schatz argument [28]. To estimate the last perturbed term in (5.35), define q_0 to be the largest integer less than or equal to $-\ln t/\ln 3$. From (5.4), it is clear that

$$|\eta_4|^{-q} < 3^{-q} < t \text{ for } q > q_0 \text{ and } q_0 \lesssim \ln k \lesssim k. \tag{6.6}$$

Define

$$\phi_0 := \begin{cases} \frac{x_1 - x}{h}, & 0 \le x \le x_1, \\ 0, & x > x_1. \end{cases}$$

Denote by $x_j = 0$ for $j < 0$ and $x_j = 1$ for $j > n$. We make use of the formulation of $u'(x)$ in (2.6) and the characterization of $u_h'(x)$ in (5.35) to obtain: For $x \in K_j$, $j = 1, 2, \cdots, n$,

$$\begin{aligned}
|u'(x) - u_h'(x)| &= \Big| \int_0^1 H(x,s) f(s) \sum_{m=0}^{n} \phi_m(s)\,\mathrm{d}s - u_h'(x) \Big| \\
&\lesssim \Big| \int_0^1 H(x,s) f(s) \phi_0(s)\,\mathrm{d}s \Big| + \sum_{m=1}^{j} \int_{x_{m-1}}^{x_{m+1}} \Big| \big(H(x,s) - \mathbf{i}\sin(mt_h^-) e^{\mathbf{i}jt_h^-} \big) f\phi_m \Big|\,\mathrm{d}s \\
&\quad + \sum_{m=j+1}^{n} \int_{x_{m-1}}^{x_{m+1}} \Big| \big(H(x,s) - \cos(jt_h^-) e^{\mathbf{i}mt_h^-} \big) f\phi_m \Big|\,\mathrm{d}s + t\,\|f\| \\
&\quad + \sum_{m=1}^{n} \int_{x_{m-1}}^{x_{m+1}} |\eta_4|^{-|j-m|}\,|f|\,\mathrm{d}s \\
&\lesssim \int_0^{x_1} |f|\,\mathrm{d}s + \sum_{m=1}^{j-2} \int_{x_{m-1}}^{x_{m+1}} \Big| \big(\mathbf{i}\sin ks\, e^{\mathbf{i}kx} - \mathbf{i}\sin(mt_h^-) e^{\mathbf{i}jt_h^-} \big) f\phi_m \Big|\,\mathrm{d}s \\
&\quad + \int_{x_{j-2}}^{x_{j+1}} |f|\,\mathrm{d}s + \sum_{m=j+1}^{n} \int_{x_{m-1}}^{x_{m+1}} \Big| \big(\cos kx\, e^{\mathbf{i}ks} - \cos(jt_h^-) e^{\mathbf{i}mt_h^-} \big) f\phi_m \Big|\,\mathrm{d}s \\
&\quad + t\,\|f\| + \int_{x_{j_1}}^{x_{j_2}} |f|\,\mathrm{d}s \\
&\lesssim \sum_{m=1}^{n} \big((m+j)\,|t_h^- - t| + t \big) (|f|, \phi_m) \\
&\quad + h^{\frac{1}{2}}\,\|f\|_{L^2([x_0,x_1]\cup[x_{j-2},x_{j+1}])} + (q_0 h)^{\frac{1}{2}}\,\|f\|_{L^2([x_{j_1},x_{j_2}])} + t\,\|f\|\,,
\end{aligned}$$

where $j_1 = \max\{j - q_0 - 1, 0\}$, $j_2 = \min\{j + q_0 + 1, n\}$ and we have used the Lagrange Mean Value Theorem to derive the last inequality. Noting that

$(m+j)\left|t_h^- - t\right| = (m+j)h\left|k_h^- - k\right| \le 2\left|k_h^- - k\right|$, the above inequality yields

$$\begin{aligned}|u'(x) - u_h'(x)| \lesssim & (t + \left|k_h^- - k\right|)\|f\| + h^{\frac{1}{2}}\|f\|_{L^2([x_0,x_1])} + h^{\frac{1}{2}}\|f\|_{L^2([x_{j-2},x_{j+1}])} \\ & + (q_0 h)^{\frac{1}{2}}\|f\|_{L^2([x_{j_1},x_{j_2}])}, \qquad \forall x \in K_j,\ j = 1, \cdots, n.\end{aligned} \tag{6.7}$$

As direct consequences of the above inequality, we have

$$\begin{aligned}\|(u-u_h)'\|_{L^2(\Omega)}^2 &\lesssim \left(t + \left|k_h^- - k\right|\right)^2\|f\|^2 + q_0^2 h^2\|f\|^2 + h\|f\|^2 \\ &\lesssim \left(t + \left|k_h^- - k\right|\right)^2\|f\|^2,\end{aligned} \tag{6.8}$$

where we have used $q_0 h \lesssim t$ (cf. (6.6)) and $h \lesssim t^2$ (since $k^2 h \gtrsim 1$) to derive the last inequality. On the other hand, from (6.7) we have

$$\begin{aligned}|[(u-u_h)']_j| &= |(u'(x_j+) - u_h'(x_j+)) - (u'(x_j-) - u_h'(x_j-))| \\ &\le |u'(x_j+) - u_h'(x_j+)| + |u'(x_j-) - u_h'(x_j-)| \\ &\lesssim (t + \left|k_h^- - k\right|)\|f\| + h^{\frac{1}{2}}\|f\|_{L^2([x_0,x_1])} + h^{\frac{1}{2}}\|f\|_{L^2([x_{j-3},x_{j+2}])} \\ &\quad + (q_0 h)^{\frac{1}{2}}\|f\|_{L^2([x_{(j-1)_1},x_{(j+1)_2}])}.\end{aligned}$$

Since $|\gamma| \le 1/6$,

$$\begin{aligned}\sum_{j=1}^{n-1}|\gamma|h\,|[(u-u_h)']_j|^2 &\lesssim (t + \left|k_h^- - k\right|)^2\|f\|^2 + q_0^2 h^2\|f\|^2 + h\|f\|^2 \\ &\lesssim \left(t + \left|k_h^- - k\right|\right)^2\|f\|^2,\end{aligned} \tag{6.9}$$

which together with (6.8) implies (6.3). By using Lemma 4.1, we can complete the proof of the theorem. ∎

Remark 6.2. (a) This theorem shows that the pollution error in H^1-norm is controlled by the phase difference $k - k_h^-$. Ihlenburg and Babuška [22, 23] obtained the same result for the FEM in the one dimensional case. Since the phase difference of the CIP-FE solution can be reduced by tuning the penalty parameter, so can its pollution error. Recently, the authors [31, 33] showed for the CIP-FEM and FEM in higher dimensions that the pollution errors in H^1-norm are of the same order as the phase difference of the FE solution. In the higher dimensional case, although the phase difference of the CIP-FE solution can still be reduced by tuning the penalty parameter, no theoretical result says that the reduced phase difference can also control the pollution error. This deserves a further investigation.

(b) The pollution effect of the CIP-FEM in one dimension can be eliminated by chosing appropriately the penalty parameters (cf. (6.5)). It is well-known that while the pollution effect in the FEM can be eliminated in the one dimensional case by a suitable modification of the discrete system that keeps the same stencil, this is no longer possible in higher dimensions (cf. [5]). With this regard we note that the stencil of the CIP-FEM ($\gamma \neq 0$) is different from that of the FEM, opening up for a possible reduction of the pollution also in higher dimensions. We refer to [34] for similar results on discontinuous Petrov-Galerkin methods.

(c) Notice that if the approximation space of the CIP-FEM is replaced by some C^1 element, say a Hermitic element space, then the penalty term J vanishes, and the CIP-FEM becomes the standard C^1 FEM. In order to eliminate the pollution error of

the C^1 FEM, one can add penalty terms on jumps of higher normal derivatives and adjust the penalty parameters.

VII. NUMERICAL EVALUATION

In this section we present two experiments, the first one is a one-dimensional model problem, which verifies the theoretical results; the second one is a two-dimensional Helmholtz problem on the unit square, which shows that, the optimal penalty parameter obtained in one-dimensional analysis defines a close to optimal choice also for simulations in higher dimensions, at least on Cartesian meshes.

A. One-dimensional problem

Throughout this subsection, we consider the BVP with constant right hand side $f(x) \equiv -1$.

The discrete wavenumber Unlike the best approximation, the CIP-FE solution is, in general, not in phase with the exact solution. Numerical tests show that the discrete solution has a phase delay with respect to the exact solution when $-\frac{1}{6} \leq \gamma < \gamma_o$ and has a phase lead with respect to the exact solution when $\gamma_o < \gamma \leq \frac{1}{6}$ which is similar to the FE solution [22]. Hence we can choose an appropriate value of the penalty parameter to eliminate the phase error. "Optimal" values of γ are those in a neighbourhood of γ_o. This is shown in Figure 2, where the real and the imaginary parts of both solutions are plotted for $k = 10$, $kh = 1$. There is no phase error for the CIP-FE solution.

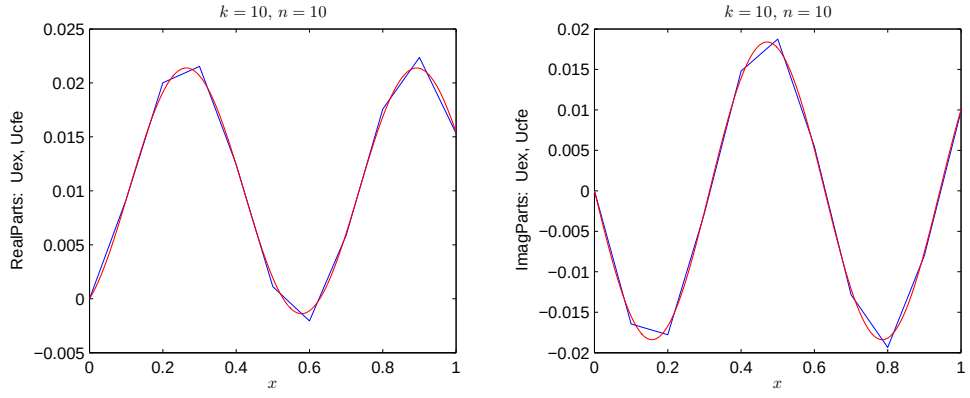

FIG. 2. No phase error of the CIP-FE solution with $\gamma = \gamma_o$ for $k = 10, n = 10$.

On a uniform mesh, the numerical dispersion relation of CIP-FE method is

$$\cos t_h^-(\gamma) = \frac{4\gamma + 1 + \frac{t^2}{6} - \sqrt{\left(1 + \frac{t^2}{6}\right)^2 + 4\gamma t^2}}{4\gamma}, \tag{7.1}$$

where $t = kh$. For fixed γ, the right-hand is a function of t, and is used for computation of the discrete wavenumber that governs the periodicity of the CIP-FE solution. In Figure 3, the functions $y_1 = \cos t = \cos t_h^-(\gamma_o)$, $y_2 = \cos t_h^-(-1/12)$, $y_3 = \cos t_h^-(0)$ and $|y_4| = 1$ are plotted. At $t_c = \sqrt{48\gamma + 12}$, the functions y_i $(i = 2, 3)$ reach absolute value 1; the numerical solution switches from the propagating case to the decaying case. The value t_c corresponds to a *cutoff frequency* for the numerical solution [30].
For fixed k, the convergence $k_h^-(:= t_h^-/h) \to k$ is visualized by $\cos t_h^- \to \cos t = \cos kh$ as $h \to 0$. The curves begin to deviate significantly at about $kh = t = 1$.

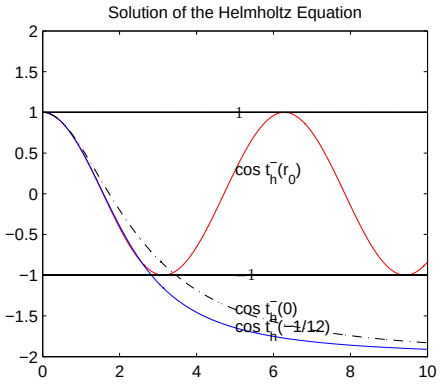

FIG. 3. Convergence of discrete to exact wavenumber via comparison of $\cos t_h^-(\gamma)$ for $\gamma = \gamma_o, -1/12, 0$ to $\cos t$. The cutoff frequency $t_c = \sqrt{8}$ for $\gamma = -1/12$, $t_c = \sqrt{12}$ for $\gamma = 0$.

Error of the best approximation and CIP-FE solution Consider in Figure 4 log-log-plots of the relative error $e_{ba} := |u - u_I|_1/|u|_1$ of the best approximation and the relative error $e_c := |u - u_h|_1/|u|_1$ by choosing $\gamma = \gamma_o$ for different k. Note that the errors first stay at 100% on coarse mesh, then start to decrease at a certain meshsize, and then decrease with constant slope of -1 (in log-log scale). This illustrates that the CIP-FE solution is convergent to the best approximation and there is no pollution error for the solution. We are interested in the critical number of DOF where the relative error begins to decrease (see for instance [22]). We can see from Figure 4 that the critical numbers of DOF for both the best approximation and the CIP-FE solution with $\gamma = \gamma_o$ are about $N = [\frac{k}{\pi}]$.
For general γ, the critical number of DOF N can be predicted using the methods of [22]:

$$|k_h^- - k| \le \frac{\pi}{3} \approx 1. \tag{7.2}$$

If γ does not depend on t, then from (7.1) and the Taylor expansion formula:

$$\begin{aligned} k_h^- h &= \cos^{-1}(\cos(t_h^-(\gamma))) \\ &= kh - \frac{12\gamma + 1}{24}(kh)^3 + \frac{1680\gamma^2 + 280\gamma + 9}{1920}(kh)^5 + O\big((kh)^7\big). \end{aligned}$$

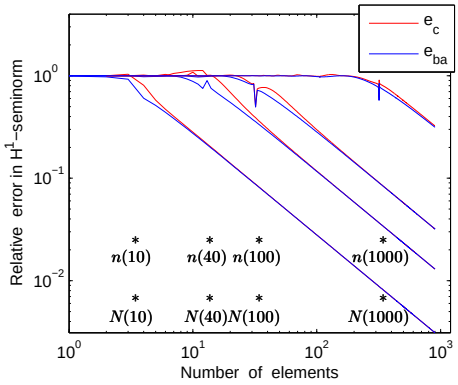

FIG. 4. The relative error of the best approximation and CIP-FE solution with $\gamma = \gamma_o$ in H^1-seminorm and predicted critical numbers of DOF for $k = 10$, $k = 40$, $k = 100$ and $k = 1000$.

Therefore, from (7.2),

$$N = \Big(\frac{|12\gamma + 1|}{24}k^3\Big)^{\frac{1}{2}} \quad (\text{if } \gamma \neq -\frac{1}{12}), \qquad N = \Big(\frac{k^5}{720}\Big)^{\frac{1}{4}} \quad (\text{if } \gamma = -\frac{1}{12}).$$

The formula of the critical number of DOF for CIP-FE solution is similar to FE solution when $\gamma \neq -1/12$, we consider the $\gamma = -1/12$ case in Figure 5. It shows that the predicted critical number of DOF is very good, especially for large k.

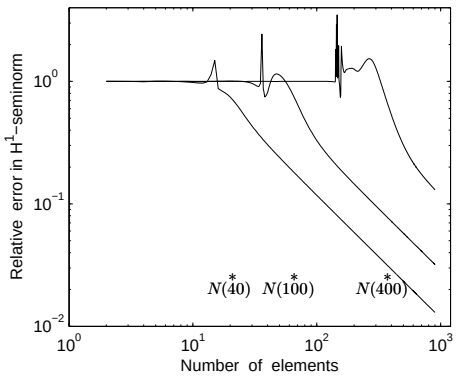

FIG. 5. The relative error of the CIP-FE solution with $\gamma = -1/12$ in H^1-seminorm and predicted critical numbers of DOF for $k = 40, 100, 400$.

Figure 6 illustrates the relative error of the CIP-FE solution for general γ other than γ_o and $-1/12$. Here we considered, $\gamma = -0.08$ and $\gamma = -0.1\mathbf{i}$, for k from 1 to 1000 on meshes determined by $k^3h^2 = 1$. It is shown that the relative error can be controlled. For small k ($1 \leq k \leq 50$), the relative error decreases rapidly with k, for large k ($k \geq 100$), the relative error is dominated by the term k^3h^2. It verifies the estimates

given by (6.3) in Theorem 6.2 and Theorem 3.4. The pollution effect does exist for the two choices of γ.

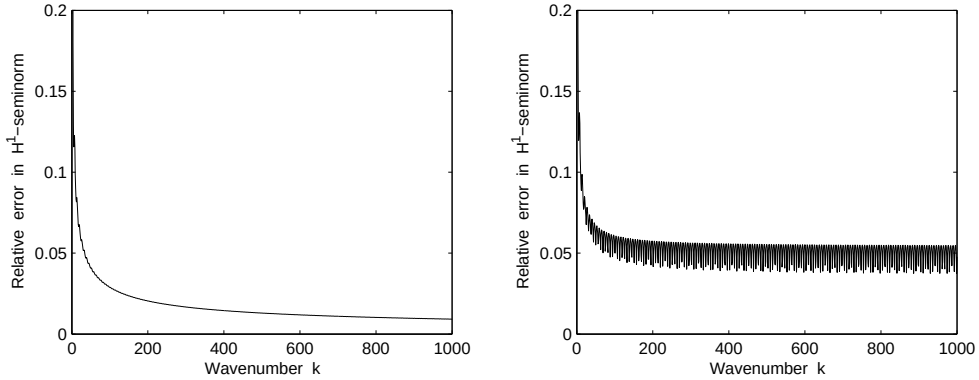

FIG. 6. Left Graph: the relative error of the CIP-FE solution with the parameter $\gamma = -0.08$ in H^1-seminorm with constraint $k^3h^2 = 1$ for k from 1 to 1000. Right graph: corresponding plot for CIP-FE solution with the parameter $\gamma = -0.1\mathbf{i}$.

In Figure 7, the ratio e_c/e_{ba} computed with the restriction $kh = 1$, is plotted for k from 1 to 1000. Obviously, the ratio (in the left of Figure 7) is increasing with k on the line. We remark that the ratio line in the right of Figure 7 is increasing with k and converges to a constant. This is due to that the relative error of the CIP-FE solution with γ (a pure imaginary number with negative imaginary part) is bounded at any range by the magnitudes of $\min\{1, k^3h^2\}$ and kh (cf. Theorem 3.4). For large k ($k \geq 100$), the ratio $e_c/e_{ba} \lesssim 1 + \min\{1, k^3h^2\}/kh = 1 + \min\{1, k\}$ (for $kh = 1$), i.e., the ratio $e_c/e_{ba} \leq C$. This shows that the imaginary part of the stabilization gives control of the amplitude of the wave.

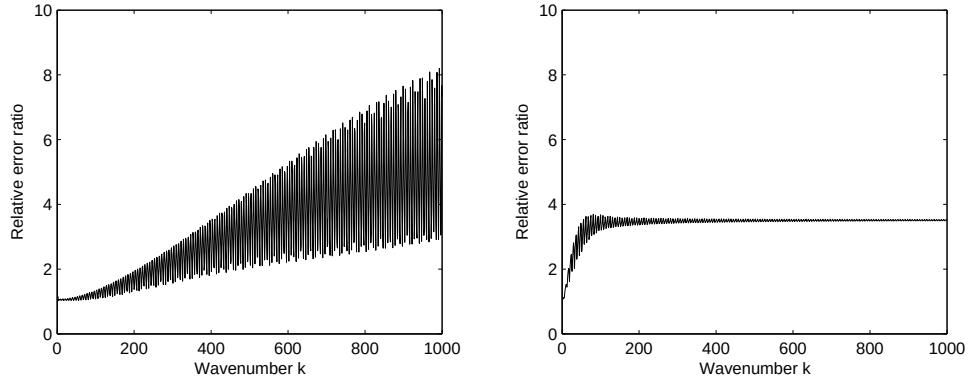

FIG. 7. Left graph: the relative error ratio e_c/e_{ba} of the CIP-FE solution with the parameter $\gamma = -0.08$ to the minimal error H^1-seminorm with constraint $kh = 1$. Right graph: corresponding plot for CIP-FE solution with the parameter $\gamma = -0.1\mathbf{i}$.

Elimination of the pollution error From Figure 6 and Figure 7, we know that the pollution error is present for general γ, but Figure 8 shows that the relative error ratio is controlled by the magnitude kh when we choose the appropriate parameter, $\gamma = \gamma_o$, for $n = k$ up to 1000. The line does neither increase nor decrease significantly with the change of k.

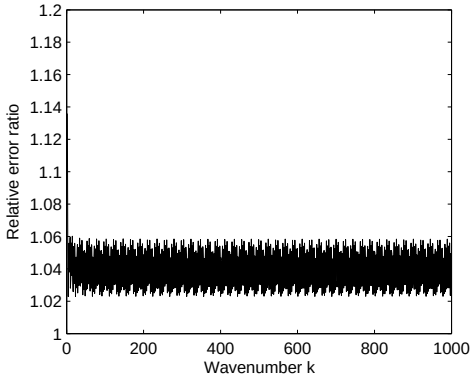

FIG. 8. The relative error ratio e_c/e_{ba} of the CIP-FE solution with $\gamma = \gamma_o$ to the minimal error H^1-seminorm with constraint $kh = 1$.

B. Two-dimensional problem

In this subsection, we show that the optimal penalty parameter γ_o given in Lemma 4.1 for the 1-d problem still works well for higher dimensional problems discretized on Cartesian meshes and the method can therefor be a cheap and efficient alternative to other approaches that are known to reduce the pollution in higher dimensions, such as those proposed in [4, 24]. We simulate the following 2-d Helmholtz problem by the bilinear FEM and CIP-FEM:

$$\begin{aligned} -\Delta u - k^2 u &= f := 0 && \text{in } \Omega, \\ \frac{\partial u}{\partial n} + \mathbf{i}ku &= g && \text{on } \Gamma, \end{aligned}$$

where Ω is the unit square $(0,1) \times (0,1)$, $\Gamma := \partial\Omega$, n denotes the unit outward normal to $\partial\Omega$ and g is so chosen that the exact solution is

$$u = J_1(kr)\sin\theta$$

in polar coordinates, where $J_\nu(z)$ are Bessel functions of the first kind. For the CIP-FEM formulation in 2-d case, we refer to [31, 33].

For any positive integer m, let $\mathcal{T}_{1/m}$ be the Cartesian grid that consists of m^2 congruent small squares of size $h = 1/m$. We remark that the number of total DOFs of both the linear FEM and CIP-FEM on $\mathcal{T}_{1/m}$ is about m^2.

Denote by $t := kh$. For the bilinear CIP-FEM, we use the penalty parameter γ_o which is obtained by a dispersion analysis for one dimensional problems such that the phase error is entirely eliminated. Although the parameter is derived for one dimensional problems, we use it in our computations for the two dimensional problem since we are

using Cartesian grids. We emphasize that for other types of meshes, e.g., triangulations, the above parameters may not be optimal. The parameters with minimum pollution effect may be found by dispersion analysis [20] or determined approximately by numerical tests [31].

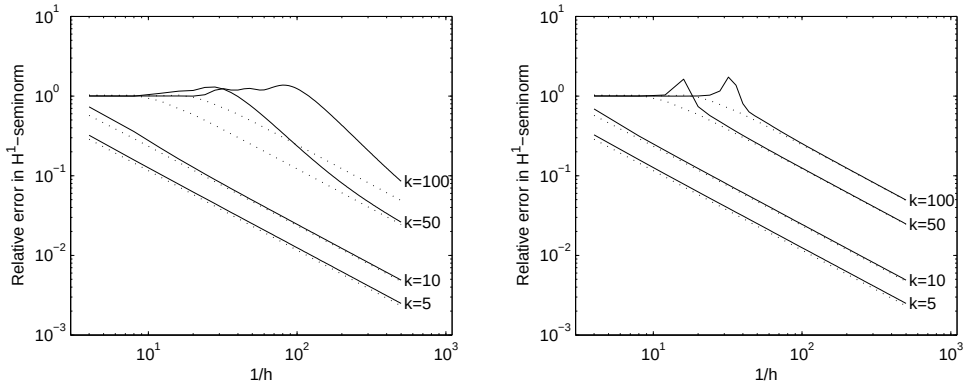

FIG. 9. Left graph: the relative error of the FE solution and the relative error of the FE interpolation (dotted) in H^1-seminorm for $k = 5, 10, 50, 100$, respectively. Right graph: corresponding plots for CIP-FE solutions with the parameter γ_o.

Figure 9 plots the relative errors in H^1-seminorm of the linear FE solutions, the linear CIP-FE solutions with penalty parameter γ_o, and the linear FE interpolations for $k = 5, 10, 50, 100$, respectively. It is shown that for $k = 5, 10$ the relative errors of both FE solutions and CIP-FE solutions fit those of the corresponding FE interpolations very well, showing that the pollution errors is insignificant for small k. For $k = 50, 100$, the relative errors of the FE solutions first stay around 100% and starts to decay only at a much higher mesh resolution compared to that of the FE interpolant. The slope is then greater than -1 in the log-log scale, indicating superconvergence in the intermediate regime. In the asymptotic regime, when hk^2 is small the convergence rate coincides with that of the FE interpolations (with slope -1). Such a behavior clearly shows the effect of pollution of the FEM for large k and h. The CIP-FE solutions behave similarly as the FE solutions but the pollution range of the former is much smaller than that of the latter, which means that the pollution effect is greatly reduced. To see this more intuitively we plot the relative errors of both methods for k from 1 to 500 with constraint $kh = 1$ in the following figure (see Figure 10). One can see that the pollution error of the linear FEM becomes dominated when k is greater than some value less than 50, while the pollution error of the linear CIP-FEM increases very slowly, which means that the pollution effect is still there but very small.

VIII. CONCLUSION

This paper provides some work for analyzing the dispersion and error of CIP-FEM. We have show the following:

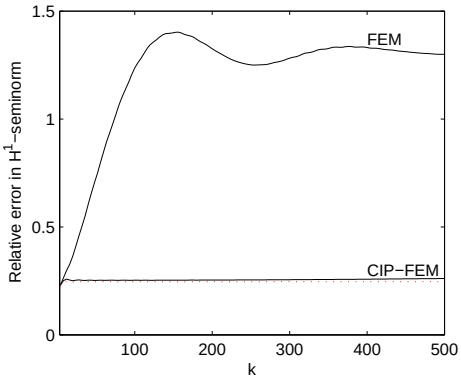

FIG. 10. The relative error of the FE solution, the CIP-FE solution with the parameter γ_o, and the FE interpolation (dashed) in H^1-seminorm with constraint $kh = 1$ for k from 1 to 500, respectively.

1. The CIP-FEM guarantees existence and amplitude control for properly chosen sign of the imaginary part of the stabilization operator.
2. There is numerical pollution for general γ and the error is mainly influenced by the pollution term for large k.
3. There are many possible "good" choices of parameters to eliminate the pollution term in one-dimensional case. Indeed, provided $kh \leq 1$ the penalty parameter may be chosen in an $O(h)$ interval of the ideal value γ_o. This together with the extended stencil give possible leads to an explanation for the success of the method in the two-dimensional case.

Future work will address the questions to what extent these results can be made to carry over to the multidimensional case and to higher polynomial orders.

REFERENCES

1. M. Ainsworth, *Discrete dispersion relation for hp-version finite element approximation at high wave number*, SIAM J. Numer. Anal., 42 (2004), pp. 553–575.
2. D. Arnold, *An interior penalty finite element method with discontinuous elements*, SIAM J. Numer. Anal., 19 (1982), pp. 742–760.
3. A.K. Aziz, R.B. Kellogg, and A.B. Stephens, *A two point boundary value problem with a rapidly oscillating solution*, Numer. Math., 53 (1998), pp. 107–121.
4. I.M. Babuška, F. Ihlenburg, E.T. Paik and S.A. Sauter, *A generalized finite element method for solving the Helmholtz equation in two dimensions with minimal pollution,* Comput. Methods Appl. Mech. Engrg. 128 (1995), 325–359.
5. I.M. Babuška and S.A. Sauter, *Is the pollution effect of the FEM avoidable for the Helmholtz equation considering high wave numbers?*, SIAM Rev., 42 (2000), pp. 451–484.
6. I. Babuška and M. Zlámal, *Nonconforming elements in the finite element method with penalty*, SIAM J. Numer. Anal., 10 (1973), pp. 863–875.

7. J. Baiges, R. Codina, *A variational multiscale method with subscales on the element boundaries for the Helmholtz equation*, Internat. J. Numer. Methods Engrg., 93 (2013), pp. 664–684.
8. G.A. Baker, *Finite element methods for elliptic equations using nonconforming elements*, Math. Comp., 31 (1977), pp. 44–59.
9. S.C. Brenner and L.R. Scott, *The mathematical theory of finite element methods*, Springer-Verlag, third ed., 2008.
10. E. Burman, *A unified analysis for conforming and nonconforming stabilized finite element methods using interior penalty*, SIAM J. Numer. Anal., 43 (2005), pp. 2012–2033.
11. E. Burman and M. Fernàndez, *Finite element methods with symmetric stabilization for the transient convection-diffusion-reaction equation*, Comput. Methods Appl. Mech. Engrg., 198 (2009), pp. 2508–2519.
12. E. Burman and P. Hansbo, *Edge stabilization for Galerkin approximations of convection-diffusion-reaction problems*, Comput. Methods Appl. Mech. Engrg., 193 (2004), pp. 1437–1453.
13. 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), pp. 2166–2188.
14. 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., 214 (2012), pp. 126–138.
15. J. Douglas Jr and T. Dupont, *Interior Penalty Procedures for Elliptic and Parabolic Galerkin methods*, Lecture Notes in Phys. 58, Springer-Verlag, Berlin, 1976.
16. J. Douglas Jr, J.E. Santos, D. Sheen, and L. Schreiyer, *Frequency domain treatment of one-dimensional scalar waves*, Math. Models Methods Appl. Sci., 3 (1993), pp. 171–194.
17. 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), pp. 782–804.
18. G. Engel, K. Garikipati, T. J. R. Hughes, M. G. Larson, L. Mazzei, and R. L. Taylor, *Continuous/discontinuous finite element approximations of fourth-order elliptic problems in structural and continuum mechanics with applications to thin beams and plates, and strain gradient elasticity*, Comput. Methods Appl. Mech. Engrg., 191 (2002), pp. 3669–3750.
19. X. Feng and H. Wu, *hp-discontinuous Galerkin methods for the Helmholtz equation with large wave number*, Math. Comp., 80 (2011), pp. 1997–2024.
20. C. Han and H. Wu, *Dispersion Analysis of the IPFEM for the Helmholtz Equation with High Wave Number on Equilateral Triangular Meshes*, Master Thesis, Nanjing University, 2015.
21. I. Harari and T.J.R. Hughes, *Finite element method for the Helmholtz equation in an exterior domain: Model problems*, Comput. Methods Appl. Mech. Engrg., 87 (1991), pp. 59–96.
22. F. Ihlenburg and I. Babuška, *Finite element solution of the Helmholtz equation with high wave number. I. The h-version of the FEM*, Comput. Math. Appl., 30 (1995), pp. 9–37.
23. ———, *Finite element solution of the Helmholtz equation with high wave number. II. The h-p version of the FEM*, SIAM J. Numer. Anal., 34 (1997), pp. 315–358.

24. A.F.D. Loula, D.T. Fernandes, *A quasi optimal Petrov-Galerkin method for Helmholtz problem,* Internat. J. Numer. Methods Engrg. 80 (2009), 1595–1622.

25. JM Melenk, A Parsania, and S Sauter, *General DG-methods for highly indefinite Helmholtz problems*, Journal of Scientific Computing, pp. 1–46.

26. 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.

27. A.A. Oberai and P.M. Pinsky, *A residual-based finite element method for the Helmholtz equation*, Internat. J. Numer. Methods Engrg., 49 (2000), pp. 399–419.

28. A.H. Schatz, *An observation concerning Ritz–Galerkin methods with indefinite bilinear forms*, Math. Comp., 28 (1974), pp. 959–962.

29. J. Shen and L.L. Wang, *Analysis of a spectral-Galerkin approximation to the Helmholtz equation in exterior domains*, SIAM J. Numer. Anal., 45 (2007), pp. 1954–1978.

30. L.L. Thompson and P.M. Pinsky, *Complex wavenumber Fourier analysis of the p-version finite element method*, Computational Mechanics, 13 (1994), pp. 255–275.

31. H. Wu, *Pre-asymptotic error analysis of CIP-FEM and FEM for Helmholtz equation with high wave number. Part I: Linear version*, IMA J. Numer. Anal., to appear. (See also arXiv: 1106.4079v1).

32. Q. Yan, *The problem of the electromagnetic scattering from a two-dimensional large open cavity*, master's thesis, Nanjing University, China, 2014.

33. L. Zhu and H. Wu, *Pre-asymptotic error analysis of CIP-FEM and FEM for Helmholtz equation with high wave number. Part II: hp version*, SIAM J. Numer. Anal., 51 (2013), pp. 1828–1852.

34. J. Zitelli, I. Muga, L. Demkowicz, J. Gopalakrishnan, D. Pardo, and V.M. Calo, *A class of discontinuous Petrov-Galerkin methods. Part IV: The optimal test norm and time-harmonic wave propagation in 1D*, J. Comput. Phys., 230 (2011), pp. 2406 – 2432.