Application of a complete radiation boundary condition for the Helmholtz equation in locally perturbed waveguides

Seungil Kim

Department of Mathematics and Research Institute for Basic Sciences, Kyung Hee University, Seoul 02447, Republic of Korea

Abstract

This paper deals with an application of a complete radiation boundary condition (CRBC) for the Helmholtz equation in locally perturbed waveguides. The CRBC, one of efficient high-order absorbing boundary conditions, has been analyzed in straight waveguides in [17]. In this paper, we apply CRBC to the Helmholtz equation posed in locally perturbed waveguides and establish the well-posedness of the problem and convergence of CRBC approximate solutions. The new CRBC proposed in this paper improves the one studied in [17] in two aspects. The first one is that the new CRBC involves more damping parameters with the same computational cost as that of CRBC in [17], which results in 50% smaller reflection errors. The second one is that the new CRBC takes a Neumann terminal condition of three term recurrence relations of auxiliary variables instead of a Dirichlet terminal condition used in [17] so that it can treat cutoff modes effectively. Finally, we present numerical experiments illustrating the convergence theory.

Keywords: Complete radiation boundary condition, absorbing boundary condition, Helmholtz equation, waveguide

1. Introduction

Waveguides are important structures used for transferring acoustic/electromagnetic energy from one point to another. In many applications, waveguides include bounded inhomogeneities such as guide vanes, fans, some obstacles or local perturbation. In this paper, we analyze an application of a high-order absorbing boundary condition, so-called complete radiation boundary condition, to locally perturbed waveguide structures including any bounded inhomogeneity.

For the numerical study of wave propagation in waveguides, it is essential to truncate the infinite waveguide into a finite domain of interest and impose

Email address: sikim@khu.ac.kr (Seungil Kim)

an accurate absorbing boundary condition on artificial boundaries. Absorbing boundary conditions, that can reduce reflected waves as small as we wish, have been studied in a great variety of different approaches based on such as perfectly matched layers (PMLs) [4, 24], truncated Dirichlet-to-Neumann (DtN) operators [6, 16, 18, 28] and rational approximations to the DtN operator [10, 12, 21] among others. Also, complete radiation boundary conditions were employed to study wave propagation in straight waveguides in [17, 23]. The CRBC is developed based on Higdon's high-order absorbing boundary condition [19, 20] but it is modified to be more suitable for numerical applications by introducing auxiliary variables satisfying certain recursive relations with some parameters, which can be tuned for minimizing reflections from absorbing boundaries.

In this paper, we extend applications of CRBC to general waveguides consisting of straight semi-infinite waveguides outside of a compact set. They include locally perturbed straight waveguides, waveguide bends, coupled waveguides and waveguides with heterogeneous inclusions. Compared with CRBC studied in [17], the new CRBC in this paper is improved in two aspects. The first improvement is that the new CRBC can allow two-sided parameter selection and hence it employs more parameters in recurrence formulas of auxiliary variables of CRBC. Consequently, it turns out that reflection errors can be made 50% smaller without increase of computational costs than those of CRBC in [17]. The second one is that the recursive formulas of auxiliary variables of CRBC are terminated with a Neumann condition instead of a Dirichlet condition used in [17]. It is found that when the Dirichlet terminal condition is used, the reflection generated by cutoff modes decreases slowly with respect to the number of parameters of CRBC. In contrast, as will be seen later, the CRBC with the Neumann terminal condition not only provides efficient treatment for all propagating modes and all important evanescent modes but also yields the exact radiation condition for cutoff modes.

In order to study CRBC applied to the Helmholtz equation for the radiation condition theoretically and numerically, we will give three different but equivalent interpretations for CRBC. The first one is the main ingredient for the theoretical part of the paper. The CRBC can be interpreted as approximate DtN operators between Sobolev spaces and this form of CRBC will be used for the well-posedness analysis of the problem with CRBC and convergence of approximate solutions. The second representation of CRBC is thought of as rational approximations to the square root function involved in the DtN operator. An optimal choice of parameters for minimizing errors in some spectral ranges is investigated from the view point of rational approximations. We note that there has been intensive research for finding rational approximations to the square root function involved in the DtN operator in terms of finite difference grid points, see e.g., [11, 12, 21]. These grid points satisfying three term recursions can be chosen in such a way that the resulting rational functions associated with the grid points minimize errors from the square root function. The third one is a practical reformulation of CRBC as studied in [17]. Due to this formulation, we can have auxiliary functions defined only on the absorbing boundaries satisfying three term recursive relations. We use this form of CRBC

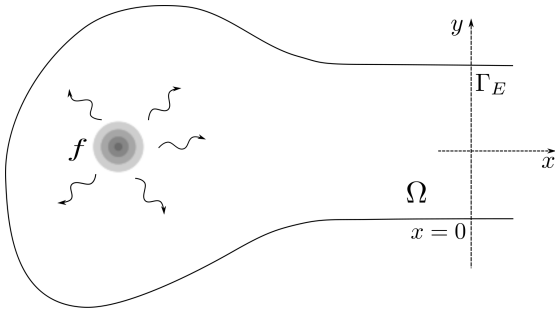


Figure 1: Domain of the model problem in $\mathbb{R}^2$

for numerical experiments.

The outline of the paper is as follows. In Section 2, we introduce CRBC with two-sided parameters and a Neumann terminal condition. Here we interpret CRBC in two different forms including an approximate DtN operator and a rational approximation of the square root function. In Section 3, we discuss selection of CRBC parameters based on the theory developed in Section 2. In Section 4, we study the well-posedness of the problem supplemented with CRBC on an artificial boundary and show the exponential convergence of approximate solutions satisfying CRBC with increasing order of CRBC. Section 5 is devoted to introducing a practical form of CRBC suitable for discretization methods. The equivalence between different forms of CRBC will be established as well. In Section 6, numerical experiments validating the convergence theory will be presented.

2. Complete radiation boundary conditions for approximate radiation conditions

Let Ω_∞ be a domain with Lipschitz boundary in $\mathbb{R}^d$ with $d = 2, 3$ obtained by merging semi-infinite waveguides such that it consists of straight semi-infinite waveguides outside of a compact set. It can be a locally perturbed waveguide, waveguide bends or waveguides with junctions of arbitrary shape. For simple presentation we assume that there exists only one semi-infinite straight waveguide in Ω_∞ with boundary $\partial\Omega_\infty$ as depicted in Figure 1, however the analysis for this simple case can be extended to the more general cases mentioned above without essential changes. Here we also assume that the axis of the waveguide Ω_∞ is parallel to the x-axis for $(x, y) \in \mathbb{R} \times \mathbb{R}^{d-1}$ and there exists a positive constant δ such that $\Omega_\infty \cap \{(x, y) \in \mathbb{R}^d \ : \ x > -\delta\}$ is a straight semi-infinite waveguide, that is, $(-\delta, \infty) \times \Theta$ with Θ a bounded domain in $\mathbb{R}^{d-1}$ with Lipschitz boundary. The model problem to be considered is to find a solution u^{ex} satisfying

$$\begin{aligned} -\Delta u^{ex} - k^2 u^{ex} &= f \text{ in } \Omega_\infty, \\ \frac{\partial u^{ex}}{\partial \nu} &= 0 \text{ on } \partial\Omega_\infty \end{aligned} \tag{2.1}$$

and a radiation condition at infinity, where f is a wave source compactly supported in $\Omega_\infty \cap \{(x,y) \in \mathbb{R}^d \ : \ x < -\delta\}$, k is a positive wavenumber and ν represents the outward unit normal vector on $\partial\Omega_\infty$. It is also possible to consider the problem with the Dirichlet condition on the boundary instead of the Neumann condition, however we focus only on this model problem as the analysis can be carried out without any significant change.

It is well-known that there may exist eigenvalues of the model problem. Analyses on discrete spectrum of waveguides can be found, for instance, for the case of waveguides with obstacles in [9, 13, 22, 26], curved waveguides in [25], locally perturbed straight waveguides in [7, 15] and coupling several waveguides with a window in [14] among others. Therefore, we need to assume that k^2 is not an eigenvalue of the model problem for the well-posedness in the infinite domain Ω_∞.

In order to introduce CRBC to the model problem, we define $\Omega = \Omega_\infty \cap \{(x,y) \in \mathbb{R}^d \ : \ x < 0\}$ for the domain of interest and denote a part of the boundary of Ω at $x = 0$ by Γ_E for the absorbing boundary and its complement from $\partial\Omega$ by Γ_c, so that $\partial\Omega = \bar{\Gamma}_E \cup \bar{\Gamma}_c$ and $\Gamma_E \cap \Gamma_c = \varnothing$ (see Figure 1). Let $\{Y_n\}_{n=0}^\infty$ be a complete set of orthonormal eigenfunctions of the negative transversal Laplace operator $-\Delta_y$ associated with eigenvalues λ_n^2 on Θ,

$$\begin{aligned} \Delta_y Y_n + \lambda_n^2 Y_n = 0 \text{ in } \Theta, \\ \frac{\partial Y_n}{\partial \nu} = 0 \text{ on } \partial\Theta. \end{aligned}$$

Here λ_n are ordered increasingly, $0 \le \lambda_0 \le \lambda_1 \le \dots$. Since eigenvalues λ_n^2 are asymptotically

$$\lambda_n^2 = O(n^{2/(d-1)}) \tag{2.2}$$

for large n, [see, e.g., 8, Ch. VI, Thm. 20-21], there exists an integer N such that $\lambda_n^2 \le k^2$ for $n \le N$ and $\lambda_n^2 > k^2$ for $n > N$. In certain situations, there may exist a positive integer $n = N$ such that $\lambda_N^2 = k^2$, which corresponds to cutoff modes. In order to understand the behavior of CRBC for all modes including propagating, evanescent and cutoff modes, we assume that such N exists. The complete orthonormal basis $\{Y_n\}_{n=0}^\infty$ allows us to express general solutions to the Helmholtz equation for $x > -\delta$ in Fourier series,

$$u(x,y) = (A_N + B_N x)Y_N(y) + \sum_{n \ne N} (A_n e^{i\mu_n x} + B_n e^{-i\mu_n x})Y_n(y) \tag{2.3}$$

with $\mu_n^2 = k^2 - \lambda_n^2$. Here μ_n are axial frequencies defined as

$$\mu_n = \begin{cases} \sqrt{k^2 - \lambda_n^2} & \text{if } 0 \le n < N, \\ 0 & \text{if } n = N, \\ i\sqrt{\lambda_n^2 - k^2} := i\tilde{\mu}_n \text{ with } \tilde{\mu}_n > 0 & \text{if } n > N. \end{cases}$$

Taking the time-harmonic dependence $e^{-i\omega t}$ into account with angular frequency $\omega = ck$ and sound speed c, the radiating solution (bounded and right-going to

infinity) is one that takes the form

$$u(x,y) = \sum_{n=0}^{\infty} A_n e^{i\mu_n x} Y_n(y).$$

We note that components for $0 \le n < N$ represent propagating modes and those for $n > N$ do evanescent modes. Thus the radiating solution can be interpreted as the solution satisfying

$$\frac{\partial u}{\partial \nu} = T(u) \text{ on } \Gamma_E,$$

where $T : H^{1/2}(\Theta) \to H^{-1/2}(\Theta)$ is the Dirichlet-to-Neumann (DtN) operator defined by

$$T(\phi) = \sum_{n=0}^{\infty} i\mu_n \phi_n Y_n$$

for $\phi = \sum_{n=0}^{\infty} \phi_n Y_n$ in $H^{1/2}(\Theta)$. Here $H^s(\Gamma_E)$ for $-1 \le s \le 1$ is a Sobolev space equipped with a norm

$$\|\phi\|^2_{H^s(\Theta)} = \sum_{n=0}^{\infty} (1+\lambda_n^2)^s |\phi_n|^2. \tag{2.4}$$

Then the radiating solution $u^{ex} \in H^1(\Omega)$ is a solution satisfying the problem

$$\begin{aligned} -\Delta u^{ex} - k^2 u^{ex} &= f \text{ in } \Omega, \\ \frac{\partial u^{ex}}{\partial \nu} &= 0 \text{ on } \Gamma_c, \\ \frac{\partial u^{ex}}{\partial \nu} - T(u^{ex}) &= 0 \text{ on } \Gamma_E. \end{aligned} \tag{2.5}$$

Here we consider the problem with wave source $f \in H^{-1}(\Omega)$ supported for $x < -\delta$, where $H^{-1}(\Omega)$ denotes the dual space of the Sobolev space $H^1(\Omega)$.

Now, we introduce CRBC that can approximate the DtN boundary condition. For positive integers n_p and n_e, CRBC of order (n_p, n_e) on Γ_E can be defined in terms of auxiliary variables φ_j satisfying the Helmholtz equation on a neighborhood of Γ_E with the homogeneous Neumann condition on the waveguide boundary and the recursive relations on a neighborhood of Γ_E,

$$\varphi_0 = u \tag{2.6}$$

$$(\frac{\partial}{\partial x} + a_j)\varphi_j = (-\frac{\partial}{\partial x} + \tilde{a}_j)\varphi_{j+1} \tag{2.7}$$

for $0 \le j \le P := n_p + n_e - 1$ together with the terminal condition

$$\frac{\partial}{\partial x}\varphi_{P+1} = 0 \text{ on } \Gamma_E, \tag{2.8}$$

where a_j and $\tilde{a}_j$ are damping parameters given by

$$a_j = \begin{cases} -ikc_j & \text{for } 0 \le j < n_p, \\ \sigma_{j-n_p} & \text{for } n_p \le j < n_p + n_e, \end{cases}$$
$$\tilde{a}_j = \begin{cases} -ik\tilde{c}_j & \text{for } 0 \le j < n_p, \\ \tilde{\sigma}_{j-n_p} & \text{for } n_p \le j < n_p + n_e \end{cases} \tag{2.9}$$

with the conditions

$$\begin{aligned} &\mu_{min}/k \le c_0 < \tilde{c}_0 < \ldots < c_{n_p-1} < \tilde{c}_{n_p-1} \le 1, \\ &\tilde{\mu}_{min} \le \sigma_0 < \tilde{\sigma}_0 < \ldots < \sigma_{n_e-1} < \tilde{\sigma}_{n_e-1} \le \tilde{\mu}_{max}. \end{aligned} \tag{2.10}$$

Here μ_{min} and $\tilde{\mu}_{min}$ are the smallest axial frequency of propagating modes and the smallest decay rate of evanescent modes, respectively, that is, if a cutoff mode exists and the corresponding eigenvalue λ_N^2 is of multiplicity 1, then

$$\mu_{min} = \mu_{N-1} \quad \text{and} \quad \tilde{\mu}_{min} = \tilde{\mu}_{N+1}.$$

Also, $\tilde{\mu}_{max}$ is determined in a way that $e^{-\delta\tilde{\mu}_n}$ for $\tilde{\mu}_n \ge \tilde{\mu}_{max}$ decays to low enough level of the reflection error of propagating modes. The optimal parameters for minimizing reflection errors resulting from the absorbing boundary condition will be discussed in Subsection 2.2.

The following subsections are devoted to interpreting CRBC in different forms.

2.1. CRBC as an approximate DtN operator

In this subsection we shall interpret CRBC as an approximate DtN operator. Since the auxiliary variables can be written as the series similar to (2.3)

$$\varphi_j(x,y) = (A_N^j + B_N^j x)Y_N(y) + \sum_{n\ne N} (A_n^j e^{i\mu_n x} + B_n^j e^{-i\mu_n x})Y_n(y),$$

the formulas (2.6)-(2.8) show that for $n \ne N$

$$(a_j + i\mu_n)A_n^j = (\tilde{a}_j - i\mu_n)A_n^{j+1} \quad \text{and} \quad (a_j - i\mu_n)B_n^j = (\tilde{a}_j + i\mu_n)B_n^{j+1} \tag{2.11}$$

with the terminal condition

$$A_n^{P+1} = B_n^{P+1}. \tag{2.12}$$

Denoting

$$R_{j,m}^n = \begin{cases} \displaystyle\prod_{\ell=j}^{m} \frac{\tilde{a}_\ell + i\mu_n}{a_\ell - i\mu_n} & \text{for } j \le m, \\ 1 & \text{for } j > m, \end{cases} \qquad Q_{j,m}^n = \begin{cases} \displaystyle\prod_{\ell=j}^{m} \frac{a_\ell + i\mu_n}{\tilde{a}_\ell - i\mu_n} & \text{for } j \le m, \\ 1 & \text{for } j > m, \end{cases}$$

it is easy to see from (2.11) and (2.12) that

$$B_n = R_{0,P}^n B_n^{P+1} \quad \text{and} \quad A_n^{P+1} = Q_{0,P}^n A_n,$$

from which we obtain

$$B_n = Z_{0,P}^n A_n. \tag{2.13}$$

Here

$$Z_{j,m}^n = Q_{j,m}^n R_{j,m}^n = \prod_{\ell=j}^{m} \frac{(a_\ell + i\mu_n)(\tilde{a}_\ell + i\mu_n)}{(a_\ell - i\mu_n)(\tilde{a}_\ell - i\mu_n)}. \tag{2.14}$$

Since $|Z_{0,P}^n|$ is the ratio of the amplitude of the left-going (reflecting) component to that of the right-going (radiating) component, it is called the reflection coefficient of the n-th mode. For $n = N$, we get $a_j B_N^j = \tilde{a}_j B_N^{j+1}$ with the terminal condition $B_N^P = 0$, which results in

$$B_N = 0. \tag{2.15}$$

Now, for given Dirichlet trace $u = \sum_{n=0}^{\infty} u_n Y_n$ of u on Γ_E, that is, $u_n = A_n + B_n = (1 + Z_{0,P}^n)A_n$ from (2.13) for $n \neq N$ and $u_N = A_N$ from (2.15) for $n = N$, the Neumann trace of u on Γ_E is given by

$$\frac{\partial u}{\partial \nu} = B_N Y_N + \sum_{n \neq N} i\mu_n (A_n - B_n) Y_n = \sum_{n=0}^{\infty} i\mu_n \frac{1 - Z_{0,P}^n}{1 + Z_{0,P}^n} u_n Y_n$$

due to (2.13) and (2.15). Finally, we define $T_c : H^{1/2}(\Gamma_E) \to H^{-1/2}(\Gamma_E)$ by

$$T_c(u) = \sum_{n=0}^{\infty} i\mu_n \left(\frac{1 - Z_{0,P}^n}{1 + Z_{0,P}^n} \right) u_n Y_n \tag{2.16}$$

for $u = \sum_{n=0}^{\infty} u_n Y_n$, which is a Dirichlet-to-Neumann operator for solutions to the Helmholtz equation satisfying CRBC on Γ_E. Here, we note that

$$|Z_{0,P}^n| = \begin{cases} \displaystyle\prod_{j=0}^{n_p-1} \left| \frac{(kc_j - \mu_n)(k\tilde{c}_j - \mu_n)}{(kc_j + \mu_n)(k\tilde{c}_j + \mu_n)} \right| < 1 & \text{for } n < N, \\ \displaystyle\prod_{j=n_p}^{n_p+n_e-1} \left| \frac{(\sigma_j - \tilde{\mu}_n)(\tilde{\sigma}_j - \tilde{\mu}_n)}{(\sigma_j + \tilde{\mu}_n)(\tilde{\sigma}_j + \tilde{\mu}_n)} \right| < 1 & \text{for } n > N. \end{cases}$$

In addition, since $\lim_{n\to\infty} Z_{0,P}^n = 1$ for given parameters a_j and $\tilde{a}_j$, $|(1 - Z_{0,P}^n)/(1 + Z_{0,P}^n)|$ is bounded independently of n and hence the operator T_c is well-defined. Thus, the problem supplemented with CRBC on Γ_E can be written in terms of the approximate DtN operator T_c,

$$\begin{aligned} -\Delta u - k^2 u &= f \text{ in } \Omega, \\ \frac{\partial u}{\partial \nu} &= 0 \text{ on } \Gamma_c, \\ \frac{\partial u}{\partial \nu} - T_c(u) &= 0 \text{ on } \Gamma_E. \end{aligned} \tag{2.17}$$

2.2. CRBC as a rational approximation

The exact DtN boundary condition for the n-th mode u_n of the solution u is given by

$$\frac{\partial u_n}{\partial \nu} = i\mu_n u_n \text{ on } \Gamma_E. \tag{2.18}$$

If we denote $-\mu_n^2 = z_n$, the DtN operator for the n-mode is the multiplication by $\sqrt{z_n} = i\mu_n$.

On the other hand, CRBC for the n-th mode is defined as

$$\frac{\partial u_n}{\partial \nu} = S(z_n)u_n, \tag{2.19}$$

where

$$S(z_n) = i\mu_n \frac{1 - Z_{0,P}^n}{1 + Z_{0,P}^n}. \tag{2.20}$$

Denoting $\mathcal{P}(x) = \prod_{j=0}^{P}(a_j + x)(\tilde{a}_j + x)$, a polynomial of degree $2P + 2$, the reflection coefficient $Z_{0,P}^n$ can be written as

$$Z_{0,P}^n = \frac{\mathcal{P}(i\mu_n)}{\mathcal{P}(-i\mu_n)}.$$

Then it is clear that

$$S(z_n) = \sqrt{z_n}\frac{\mathcal{P}(-\sqrt{z_n}) - \mathcal{P}(\sqrt{z_n})}{\mathcal{P}(-\sqrt{z_n}) + \mathcal{P}(\sqrt{z_n})}$$

is an even rational approximation to $\sqrt{z_n}$ of $(2P+2, 2P+2)$-type and so $S(z_n)$ can be viewed as a rational approximation of $(P+1, P+1)$-type as a function of z_n. Since the relative averaged error of the rational approximation is given by

$$2\left|\frac{\sqrt{z_n} - S(z_n)}{\sqrt{z_n} + S(z_n)}\right| = 2\left|Z_{0,P}^n\right|,$$

the convergence of the rational approximation depends considerably on selecting parameters under the conditions (2.9) with regard to the minimization of the reflection coefficient $|Z_{0,P}^n|$. Alternatively, we may select parameters a_j and $\tilde{a}_j$ in $\mathbb{C}$ rather than real or purely imaginary parameters to minimize the error of $S(z)$ over $[-k^2, -\mu_{min}^2] \cup [\tilde{\mu}_{min}^2, \tilde{\mu}_{max}^2]$. However since it is much easier to handle the minimization problem with parameters satisfying (2.9) than the general minimization problem in the complex plane, we restrict ourselves to the minimization problem with the conditions (2.9), and it allows us to separate the minimization of $|Z_{0,P}^n|$ for propagating modes and evanescent modes. Here we note that when the parameters are chosen to be purely imaginary for propagating modes and real for evanescent modes, this problem is related with the third Zolotarev problem, see e.g., [29].

Noting that

$$\left|\frac{(a_j + i\mu_n)(\tilde{a}_j + i\mu_n)}{(a_j - i\mu_n)(\tilde{a}_j - i\mu_n)}\right| = 1 \text{ for } \begin{cases} n_p \le j < n_p + n_e & \text{if } 0 \le n \le N, \\ 0 \le j < n_p & \text{if } N \le n, \end{cases} \tag{2.21}$$

the optimal parameters a_j and $\tilde{a}_j$ can be obtained by solving two separate min-max problems

$$\rho_p \equiv \min_{\substack{a_0,\dots,a_{n_p-1},\\ \tilde{a}_0,\dots,\tilde{a}_{n_p-1}\in i\mathbb{R}_-}} \max_{\mu\in[\mu_{min},k]} \prod_{j=0}^{n_p-1} \left|\frac{(a_j+i\mu)(\tilde{a}_j+i\mu)}{(a_j-i\mu)(\tilde{a}_j-i\mu)}\right|, \tag{2.22}$$

$$\rho_e \equiv \min_{\substack{a_{n_p},\dots,a_{n_p+n_e-1}\\ \tilde{a}_{n_p},\dots,\tilde{a}_{n_p+n_e-1}\in\mathbb{R}_+}} \max_{\mu\in[\tilde{\mu}_{min},\tilde{\mu}_{max}]} \prod_{j=n_p}^{n_p+n_e-1} \left|\frac{(a_j-\mu)(\tilde{a}_j-\mu)}{(a_j+\mu)(\tilde{a}_j+\mu)}\right|. \tag{2.23}$$

In [2, 27] the analytic solutions to the min-max problems are investigated by using elliptic functions. The exponential decay of the reflection coefficients is established as well. For more precise discussion, we recall the elliptic functions defined in terms of the complete elliptic integral of the first kind

$$K(\gamma) = \int_0^1 \frac{1}{\sqrt{(1-t^2)(1-\gamma^2t^2)}}dt$$

for $0<\gamma<1$: the sine amplitude $\mathrm{sn}(w,\gamma)$ is a map from the interval $[0,K(\gamma)]$ to the another interval $[0,1]$ given by the relation

$$w = \int_0^{\mathrm{sn}(w,\gamma)} \frac{1}{\sqrt{(1-t^2)(1-\gamma^2t^2)}}dt, \tag{2.24}$$

and the delta amplitude $\mathrm{dn}(w,\eta)$ is defined by

$$\mathrm{dn}(w,\gamma) = \sqrt{1-\gamma^2\mathrm{sn}^2(w,\gamma)}.$$

Let $\gamma_p = \mu_{min}/k$ and $\gamma_e = \tilde{\mu}_{min}/\tilde{\mu}_{max}$ and denote $\tilde{\gamma}_p = \sqrt{1-\gamma_p^2}$ and $\tilde{\gamma}_e = \sqrt{1-\gamma_e^2}$. We use $(\rho,\gamma,\tilde{\gamma},m) = (\rho_p,\gamma_p,\tilde{\gamma}_p,2n_p)$ for the propagating component and $(\rho,\gamma,\tilde{\gamma},m) = (\rho_e,\gamma_e,\tilde{\gamma}_e,2n_e)$ for the evanescent component for the sake of simple presentation. Then the CRBC parameters are related with the solutions to the Zolotarev problem on the intervals $[\gamma,1]$, that can be given by

$$s_j = \mathrm{dn}\left(\left(1-\frac{2j+1}{2m}\right)K(\tilde{\gamma}),\tilde{\gamma}\right) \text{ for } j=0,\dots,m-1 \tag{2.25}$$

as seen in Appendix A of [10]. See also [2, 5, 27]. With these values, we obtain that

$$\begin{array}{llll} c_j = s_{2j}, & \tilde{c}_j = s_{2j+1} & \text{for } \gamma=\gamma_p \\ \sigma_j = s_{2j}\tilde{\mu}_{max}, & \tilde{\sigma}_j = s_{2j+1}\tilde{\mu}_{max} & \text{for } \gamma=\gamma_e. \end{array} \tag{2.26}$$

It is also shown in [10, A.2] or [27, (3.17)] that the reflection coefficient ρ satisfies

$$\frac{2q^{m/4}}{1+q^{m/2}} \le \rho \le 2q^{m/4}, \tag{2.27}$$

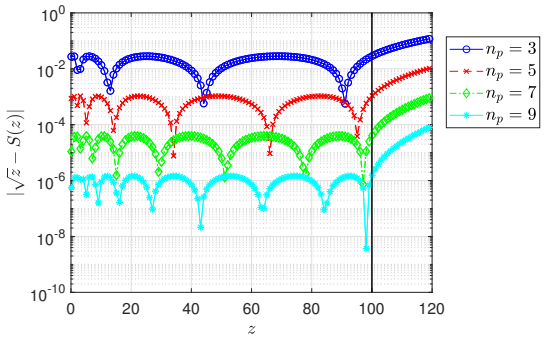


Figure 2: Rational approximation $S(z)$ to $\sqrt{z}$ with optimal parameters for $n_p = 3, 5, 7, 9$.

with Jacobi's nome $q = \exp(-\pi K(\tilde{\eta})/K(\eta))$, where η and $\tilde{\eta}$ are given by $\eta^{1/2} = (1-\gamma^{1/2})/(1+\gamma^{1/2}), \tilde{\eta} = \sqrt{1-\eta^2}$. In addition, the approximate behavior of ρ,

$$\rho \leq e^{-Cm/\ln(1/\gamma)}, \tag{2.28}$$

can be shown by using an approximate expression of the nome as seen in [21, A.9] (see also [29]). Hereafter C denotes a generic constant that may take different values at different places but does not depend on n_p, n_e and CRBC parameters. In order to demonstrate the convergence of the rational approximation $S(z)$ to $\sqrt{z}$ in the interval $[0.01, 100]$ as an example, we choose bounds for parameters to be $\mu_{min} = 0.1$, $k = 10$ and determine optimal parameters for $n_p = 3, 5, 7, 9$. The results are presented in Figure 2, where we observe that the errors in $S(z)$ are reduced uniformly in the interval $[0.01, 100]$ as n_p increases.

Here we note that compared with the reflection coefficients of the CRBC with $a_j = \tilde{a}_j$ analyzed in [17] which are the optimal values of the min-max problem

$$\min_{a_0,\ldots,a_{n_p-1}\in i\mathbb{R}_-} \max_{\mu\in[\mu_{min},k]} \prod_{j=0}^{n_p-1} \left| \frac{(a_j+i\mu)^2}{(a_j-i\mu)^2} \right|,$$

for instance, for propagating components, those of the new CRBC with more degrees of freedom are reduced by factor $1/2$ due to (2.27) as seen Table 1. Indeed, when n_p auxiliary variables are used for propagating components in two different CRBCs, for small $q^{n_p/2}$ the former one is approximately $\left(2q^{n_p/4}\right)^2$ whereas the latter one is given by approximately $2q^{2n_p/4}$.

3. Parameter selections

It is important to choose appropriate parameters for efficient performance of CRBC. The general idea of how to select parameters is discussed in [17]. In practice, if the number of important modes involved in solutions is small and the

k	one-sided, $a_j = \tilde{a}_j$	two-sided, $a_j \neq \tilde{a}_j$	ratio
4	1.806194e-07	9.030969e-08	0.500000
5	3.793422e-09	1.896711e-09	0.500000
6	2.571956e-10	1.285978e-10	0.500000
7	4.247227e-06	2.123613e-06	0.500000
8	1.806194e-07	9.030969e-08	0.500000
9	2.093552e-08	1.046776e-08	0.500000
10	2.288380e-05	1.144190e-05	0.500000
11	1.220426e-06	6.102130e-07	0.500000
12	1.806194e-07	9.030969e-08	0.500000
13	7.842724e-05	3.921362e-05	0.500000

Table 1: The comparison between the reflection coefficients ρ_p of the CRBC of [17] and the new CRBC for $n_p = 3$ and different wavenumbers

corresponding $-i\mu_n$ can be estimated by using a Lanczos algorithm for the cross-sectional Laplace operator, then we can use these values for parameters so that the reflection coefficients $|Z^n_{0,P}|$ for those modes vanish due to the formula (2.14), i.e., the CRBC with these parameters serves as an exact radiation condition for those modes. If it appears not to be efficient to estimate all $-i\mu_n$ for important modes, then we estimate μ_{min} and $\tilde{\mu}_{min}$ by computing eigenvalues near to k^2 and use them for the bounds of the min-max problem. In order to balance the reflection errors of three parts, we choose parameters as follows.

i. For given $n_p \geq 1$, we solve the min-max problem (2.22) to find $2n_p$ imaginary parameters and estimate ρ_p in the interval $[\mu_{min}, k]$ with $\mu_{min} = \mu_{N-1}$.

ii. We decide the upper bound $\tilde{\mu}_{max}$ of the min-max problem (2.23) for relatively slowly decaying evanescent modes by solving the inequality

$$e^{-\tilde{\mu}_{max}\delta} \leq \rho_p.$$

It guarantees that reflection errors of all evanescent modes whose decay rate $\tilde{\mu}_n$ is larger than $\tilde{\mu}_{max}$ are less than ρ_p.

iii. We decide $n_e \geq 1$ and $2n_e$ real parameters of the solution to the min-max problem (2.23) in the interval $[\tilde{\mu}_{min}, \tilde{\mu}_{max}]$ with $\tilde{\mu}_{min} = \tilde{\mu}_{N+1}$ such that

$$e^{-\tilde{\mu}_{N+1}\delta}\rho_e \leq \rho_p.$$

In this procedure, with decreasing tolerance ε the required number of parameters and $\tilde{\mu}_{max}$ satisfies

$$n_p \propto \ln\left(\frac{1}{\varepsilon}\right) \cdot \ln\left(\frac{1}{\gamma_p}\right) := \zeta_p, \tag{3.1}$$

$$\tilde{\mu}_{max} \propto \frac{1}{\delta} \cdot \ln\frac{1}{\varepsilon}, \tag{3.2}$$

$$n_e \propto \ln\left(\frac{1}{\varepsilon}\right) \cdot \ln\left(\frac{1}{\delta\tilde{\mu}_{min}}\right) + \ln\left(\frac{1}{\varepsilon}\right) \cdot \ln\ln\left(\frac{1}{\varepsilon}\right) := \zeta_e. \tag{3.3}$$

n_p	ρ_p	$\tilde{\mu}_{max}$	n_e	$e^{-\tilde{\mu}_{N+1}\delta}\rho_e$
1	2.0952e-02	7.7310e+01	2	1.5324e-03
2	2.1949e-04	1.6848e+02	4	3.3768e-05
3	2.2994e-06	2.5966e+02	6	9.4755e-07
4	2.4089e-08	3.5083e+02	9	3.6646e-09
5	2.5235e-10	4.4200e+02	11	1.5373e-10
6	2.6437e-12	5.3318e+02	14	9.5911e-13

Table 2: Reflection errors of CRBC when $k = 10\pi$, $\delta = 0.05$, $\mu_{min} = \mu_{N-1} \approx 13.6938$ and $\tilde{\mu}_{min} = \tilde{\mu}_{N+1} \approx 14.3965$ with $N = 10$.

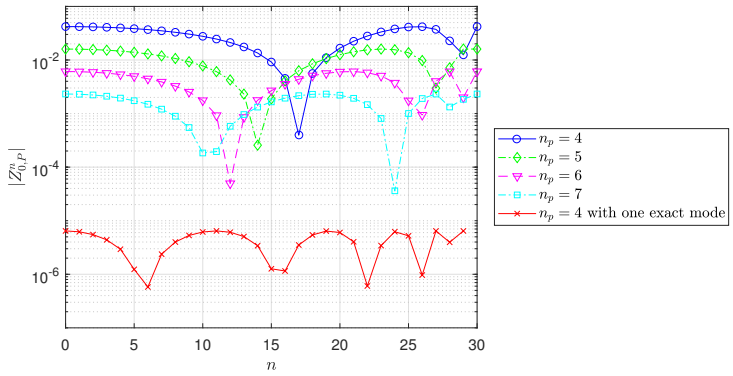


Figure 3: Reflection coefficients of propagating modes when $\mu_{min} \ll 1$.

For example, for $k = 10\pi$ and $\delta = 0.05$, CRBC can suppress the reflection errors with respect to increasing n_p as in Table 2.

In a certain situation, μ_{min} or $\tilde{\mu}_{min}$ is very small and in such a case the decay rate of ρ_p or ρ_e becomes worse as μ_{min} or $\tilde{\mu}_{min}$ is closer to zero. In particular, when there is no cutoff mode, that is, there is no λ_n such that $k \neq \lambda_n$, but k is close to λ_n so that μ_{min}(or $\tilde{\mu}_{min}$) $\ll 1$ (the mode corresponding to such μ_n is called a near-cutoff mode), we may need to increase n_p(or n_e) to achieve desirable accuracy of the absorbing boundary condition. However, instead of increasing n_p or n_e, it is preferable to use $-i\mu_n$ of near-cutoff modes for a parameter a_j and to estimate one more μ_n of the next one to the near-cutoff mode and employ it for a lower bound of the min-max problem. An example for $k = 30\pi + 10^{-6}$ and $\Theta = (0, 1)$ is illustrated in Figure 3 showing that the result of $n_p = 4$ (red-solid line) with the reflection of the near-cutoff mode ($n = 30$) eliminated by CRBC is superior to those of other $n_p = 4, 5, 6, 7$ with μ_{min} equal to μ_n of the near-cutoff mode. Thus, when we choose parameters, we first examine whether a near-cutoff mode exists. If it exists, we choose a parameter in a way that CRBC eliminates the reflection of the near-cutoff mode and follow the general rule for the rest of parameters as above. We would

like to mention that the well-known absorbing boundary condition PML suffers from reducing reflection errors of near-cutoff modes. According to the results in [24], it turns out that reflection errors of the PML method decrease at the rate of $e^{-2\sigma L}$, where L and σ stand for the PML width and strength, respectively. In addition, the number of grid points, $\mathcal{N}_{\rm PML}$, along the axis of waveguide in the PML damping zone is proportional to $\sigma L/\mu_{min}$. As a consequence, in order to keep the error less than tolerance ε we have

$$\mathcal{N}_{\rm PML} \propto \frac{1}{\mu_{min}} \ln \frac{1}{\varepsilon}.$$

Noting that $\mathcal{N}_{\rm PML}$ corresponds to the number of auxiliary variables of CRBC, PML requires extremely larger computational resource than CRBC.

In the rest of this section we give some properties of the reflection coefficient when the number of parameters solving the min-max problems grows at a rate of (3.1)-(3.3) with respect to tolerance ε.

Lemma 3.1. *Let s_j be the values defined by* (2.25)*. Then the arithmetic mean s_{av} of s_j is asymptotically equivalent to $\pi/(2K(\tilde{\gamma}))$ as $m \to \infty$, that is*

$$\lim_{m\to\infty} \frac{s_{av}}{\pi/(2K(\tilde{\gamma}))} = 1. \tag{3.4}$$

We denote it by $s_{av} \sim \pi/K(\tilde{\gamma})$ as $m \to \infty$.

Proof. Noting that $|\mathrm{dn}''(\cdot,\tilde{\gamma})| \le \tilde{\gamma} \le 1$, we can show that the midpoint quadrature rule for $\mathrm{dn}(\cdot,\tilde{\gamma})$ in the interval $[0, K(\tilde{\gamma})]$ satisfies

$$\left| \int_0^{K(\tilde{\gamma})} \mathrm{dn}(x,\tilde{\gamma})dx - \sum_{j=0}^{m-1} s_j \frac{K(\tilde{\gamma})}{m} \right| \le \frac{K(\tilde{\gamma})}{6} \left(\frac{K(\tilde{\gamma})}{2m} \right)^2.$$

Therefore, by invoking that an anti-derivative of $\mathrm{dn}(x,\tilde{\gamma})$ with respect to x is $\sin^{-1}\mathrm{sn}(x,\tilde{\gamma})$ (see [1, 16.24.3]) together with the fact that $\mathrm{sn}(K(\tilde{\gamma}),\tilde{\gamma}) = 1$ and $\mathrm{sn}(0,\tilde{\gamma}) = 0$ resulting from (2.24), we have

$$\left| \frac{\pi}{2K(\tilde{\gamma})} - s_{av} \right| \le \frac{1}{24} \left(\frac{K(\tilde{\gamma})}{m} \right)^2, \tag{3.5}$$

which can be written as

$$\left| 1 - \frac{s_{av}}{\pi/(2K(\tilde{\gamma}))} \right| \le \frac{1}{12\pi} \frac{K(\tilde{\gamma})^3}{m^2}.$$

If $\gamma = \gamma_p$, then $K(\tilde{\gamma})$ is constant and hence (3.4) follows.

In case that $\gamma = \gamma_e$, using the asymptotic behavior of $K(\tilde{\gamma}_e)$ as $\tilde{\gamma}_e \to 1$,

$$K(\tilde{\gamma}_e) \sim -\frac{1}{2}\ln(1-\tilde{\gamma}_e) \sim \ln\left(\frac{1}{\gamma_e}\right) \quad \text{as } \tilde{\gamma}_e \to 1 \tag{3.6}$$

together with (3.2) and (3.3), we obtain that

$$\lim_{m\to\infty}\frac{K(\tilde{\gamma})^3}{m^2}\le\lim_{m\to\infty}C\frac{(\ln(\tilde{\mu}_{max}/\tilde{\mu}_{min}))^3}{(\ln(1/\varepsilon)\cdot\ln\ln(1/\varepsilon))^2}=\lim_{m\to\infty}C\frac{(\ln\ln(1/\varepsilon))^3}{(\ln(1/\varepsilon)\cdot\ln\ln(1/\varepsilon))^2}=0$$

since $\varepsilon\to 0$ as $m\to\infty$, which completes the proof for $\gamma=\gamma_e$. □

Lemma 3.2. *For given $M>N$, let σ_{av} be the arithmetic mean of the real parameters $\sigma_j,\tilde{\sigma}_j$ for $0\le j<n_e$ satisfying (2.10) with $\tilde{\mu}_{max}=\tilde{\mu}_M$. Then the reflection coefficients of modes for $n>M$ are bounded by*

$$|Z^n_{0,P}|\le\left|\frac{\tilde{\mu}_n-\sigma_{av}}{\tilde{\mu}_n+\sigma_{av}}\right|^{2n_e}. \tag{3.7}$$

Proof. Using (2.21), we see that

$$|Z^n_{0,P}|=\prod_{j=0}^{n_e-1}\frac{(\tilde{\mu}_n-\sigma_j)(\tilde{\mu}_n-\tilde{\sigma}_j)}{(\tilde{\mu}_n+\sigma_j)(\tilde{\mu}_n+\tilde{\sigma}_j)}\text{ for }n>M.$$

Thus, it suffices to show that for any m positive real numbers $\alpha_1,\dots,\alpha_m$ such that $\tilde{\mu}_{min}\le\alpha_j\le\tilde{\mu}_{max}$ and for the fractional function $h(x)=(\tilde{\mu}-x)/(\tilde{\mu}+x)$ with $\tilde{\mu}>\tilde{\mu}_{max}$

$$\prod_{j=1}^m h(\alpha_j)\le\left(h\left(\frac{\sum_{j=1}^m\alpha_n}{m}\right)\right)^m.$$

Since $0<h(x)<1$ for $\tilde{\mu}_{min}\le x\le\tilde{\mu}_{max}$, the above inequality is equivalent to

$$-\ln h\left(\frac{\sum_{j=1}^n\alpha_n}{m}\right)\le-\frac{1}{m}\sum_{j=1}^m\ln h(\alpha_j). \tag{3.8}$$

Now, noting that $-\ln h(x)$ is convex in the interval $\tilde{\mu}_{min}\le x\le\tilde{\mu}_{max}$, the inequality (3.8) immediately follows from Jensen's inequality.

□

Lemma 3.3. *Assume that $M>N$ and $n_e\ge 1$ are determined by the parameter selection procedure as above for $n_p\ge 1$ so that (3.1)-(3.3) hold, and let a_j and $\tilde{a}_j$ be the parameters satisfying the min-max problems (2.22) and (2.23) with the conditions (2.10) and $\tilde{\mu}_{max}=\tilde{\mu}_M$. Then the reflection coefficients of modes for $n>M$ satisfy*

$$\frac{1}{|1+Z^n_{0,P}|}\le C. \tag{3.9}$$

Proof. We will show that $Z^n_{0,P}$ is away from -1 for all $n\ge M$ independently of n_p,n_e and CRBC parameters. In the proof, we keep in mind that $n_p\to\infty$, $n_e\to\infty$ as $\varepsilon\to 0$ and vice versa according to (3.1), (3.2) and (3.3).

We first show that for sufficiently large $\tilde{\mu}_n$

$$\arg(Z^n_{0,n_p-1})<\frac{2\pi}{3}. \tag{3.10}$$

Recalling the definition of the imaginary parameters $a_j = -ikc_j$ for $0 \le j < n_p$, since

$$0 < \arg(\tilde{\mu}_n + ikc_j) = \tan^{-1}\left(\frac{kc_j}{\tilde{\mu}_n}\right) < \frac{kc_j}{\tilde{\mu}_n} \quad \text{for } n > M$$

and the analogous inequality for $\tilde{a}_j$ holds, by using (3.5) we can show that

$$0 < \arg(Z^n_{0,n_p-1}) < \sum_{j=0}^{n_p-1} \frac{2k(c_j + \tilde{c}_j)}{\tilde{\mu}_n} \le \frac{4n_p k}{\tilde{\mu}_n}\left(\frac{\pi}{2K(\tilde{\gamma}_p)} + \frac{K(\tilde{\gamma}_p)^2}{96n_p^2}\right) \quad \text{for } n > M.$$

We denote by

$$\tilde{\mu}_{ang} := \frac{4n_p k}{2\pi/3}\left(\frac{\pi}{2K(\tilde{\gamma}_p)} + \frac{K(\tilde{\gamma}_p)^2}{96n_p^2}\right),$$

which allows us to have that if $\tilde{\mu}_n \ge \tilde{\mu}_{ang}$, then $\Re(Z^n_{0,P}) > -1/2$ and hence $|1 + Z^n_{0,P}| > 1/2$. From (3.1) we note that $\tilde{\mu}_{ang}$ has the asymptotically same order of magnitude as

$$\tilde{\mu}_{ang} \asymp \frac{3k}{K(\tilde{\gamma}_p)} \ln\left(\frac{1}{\gamma_p}\right) \cdot \ln\left(\frac{1}{\varepsilon}\right) \quad \text{as } \varepsilon \to 0 \tag{3.11}$$

with respect to tolerance ε for reflection errors.

Second, for $0 < R < 1$ we examine an upper bound $\tilde{\mu}_R$ of $\tilde{\mu}_n$ such that $|Z^n_{0,P}| < R$. We begin by solving the equation

$$\left(\frac{\tilde{\mu}_R - \sigma_{av}}{\tilde{\mu}_R + \sigma_{av}}\right)^{2n_e} = R, \tag{3.12}$$

which leads us to

$$\tilde{\mu}_R = \sigma_{av}\frac{1 + R^{1/(2n_e)}}{1 - R^{1/(2n_e)}} \sim \frac{\pi\tilde{\mu}_{max}}{2K(\tilde{\gamma}_e)}\frac{4n_e}{\ln(1/R)} \quad \text{as } n_e \to \infty.$$

Here we also used the asymptotic equivalence in Lemma 3.1. We further show by (3.6), (3.2) and (3.3) that

$$\tilde{\mu}_R \sim \frac{\pi\tilde{\mu}_{max}}{\ln(1/\gamma_e)}\frac{2n_e}{\ln(1/R)} \asymp \frac{2\pi}{\delta\ln(1/R)} \cdot \left(\ln\frac{1}{\varepsilon}\right)^2 \quad \text{as } \varepsilon \to 0. \tag{3.13}$$

Due to (3.11) and (3.13) we can show that there exists $\varepsilon_0 > 0$ such that $\tilde{\mu}_R > \tilde{\mu}_{ang}$ for $\varepsilon < \varepsilon_0$. It implies that if CRBC with large n_p and n_e is applied so that $\varepsilon < \varepsilon_0$, then $|1+Z^n_{0,P}| > 1-R$ for $\tilde{\mu}_n \le \tilde{\mu}_R$ due to (3.12) and $|1+Z^n_{0,P}| > 1/2$ for $\tilde{\mu}_n > \tilde{\mu}_R (> \tilde{\mu}_{ang})$ due to (3.10). In order to examine the case that $\varepsilon \ge \varepsilon_0$, we note that there are only finite number of CRBCs of order (n_p, n_e) that result in $\varepsilon \ge \varepsilon_0$. Thus, we can find $\mathcal{R} = \min_{\varepsilon > \varepsilon_0, n > M} |1 + Z^n_{0,P}| > 0$ independent of n_p and n_e. As a conclusion, we show that $|1 + Z^n_{0,P}| > \min\{\mathcal{R}, 1 - \max\{R, 1/2\}\}$ for $n > M$ and for any $n_p, n_e \ge 1$, which completes the proof. □

4. Well-posedness and convergence analysis

This section is devoted to establishing the well-posedness of the wave propagation problem (2.17) with CRBC for the radiation condition on Γ_E and providing the convergence analysis of approximate solutions satisfying CRBC.

4.1. Preliminaries

We consider the model problem in the truncated domain supplemented with the DtN boundary condition for a radiation condition on the artificial boundary Γ_E, assuming that k^2 is not an eigenvalue of the problem (2.1) and there are wave sources $f \in H^{-1}(\Omega)$ supported for $x < -\delta$ and $g \in H^{-1/2}(\Gamma_E)$.

$$\begin{aligned} -\Delta u - k^2 u &= f \text{ in } \Omega, \\ \frac{\partial u}{\partial \nu} &= 0 \text{ on } \Gamma_c, \\ \frac{\partial u}{\partial \nu} - T(u) &= g \text{ on } \Gamma_E. \end{aligned} \tag{4.1}$$

The corresponding variational problem is to find $u \in H^1(\Omega)$ such that

$$A(u, \phi) = \langle f, \phi \rangle_{1,\Omega} + \langle g, \phi \rangle_{\frac{1}{2},\Gamma_E} \text{ for all } \phi \in H^1(\Omega), \tag{4.2}$$

where $A(\cdot, \cdot)$ is a sesquilinear form defined in $H^1(\Omega) \times H^1(\Omega)$ by

$$A(u, \phi) = (\nabla u, \nabla \phi)_\Omega - k^2 (u, \phi)_\Omega - \langle Tu, \phi \rangle_{\frac{1}{2},\Gamma_E}.$$

From here on we use $(\cdot, \cdot)_\Omega$ and $(\cdot, \cdot)_{\Gamma_E}$ for the L^2-inner products in Ω and Γ_E, respectively. Also, $\langle \cdot, \cdot \rangle_{1,\Omega}$ and $\langle \cdot, \cdot \rangle_{\frac{1}{2},\Gamma_E}$ stand for the duality pairings between $H^{-1}(\Omega)$ and $H^1(\Omega)$ and between $H^{-1/2}(\Gamma_E)$ and $H^{1/2}(\Gamma_E)$, respectively. We have the well-posedness of the problem (4.2) as seen in [6, 16, 24].

Lemma 4.1. *Assume that k^2 is not an eigenvalue of the problem* (2.1)*. Then there exists a positive constant C such that*

$$\|u\|_{H^1(\Omega)} \le C \sup_{0 \neq \phi \in H^1(\Omega)} \frac{|A(u, \phi)|}{\|\phi\|_{H^1(\Omega)}}$$

and

$$\|u\|_{H^1(\Omega)} \le C \sup_{0 \neq \phi \in H^1(\Omega)} \frac{|A(\phi, u)|}{\|\phi\|_{H^1(\Omega)}}$$

for all $u \in H^1(\Omega)$. Therefore, the problem (4.2) *has a unique solution $u \in H^1(\Omega)$ satisfying*

$$\|u\|_{H^1(\Omega)} \le C(\|f\|_{H^{-1}(\Omega)} + \|g\|_{H^{-1/2}(\Gamma_E)}). \tag{4.3}$$

Lemma 4.1 also implies that the adjoint problem for (4.2),

$$A(\phi, u) = \langle \phi, f \rangle_{1,\Omega} + \langle \phi, g \rangle_{\frac{1}{2},\Gamma_E} \text{ for all } \phi \in H^1(\Omega)$$

has a unique solution u satisfying the stability estimate (4.3). Here we note that the solution u to the adjoint problem satisfies the boundary condition

$$\frac{\partial u}{\partial \nu} - T^*(u) = g \text{ on } \Gamma_E,$$

where $T^* : H^{1/2}(\Gamma_E) \to H^{-1/2}(\Gamma_E)$ is the adjoint operator of the DtN operator T defined by

$$\langle T(\phi), \psi \rangle_{\frac{1}{2},\Gamma_E} = \langle \phi, T^*(\psi) \rangle_{\frac{1}{2},\Gamma_E}$$

for $\phi, \psi \in H^{1/2}(\Gamma_E)$, that is, for $\phi = \sum_{n=0}^{\infty} \phi_n Y_n$

$$T^*(\phi) = \sum_{n=0}^{\infty} \overline{i\mu_n} \phi_n Y_n.$$

As an approximate adjoint DtN operator, we introduce the adjoint CRBC operator $T_c^* : H^{1/2}(\Gamma_E) \to H^{-1/2}(\Gamma_E)$ defined by

$$\langle T_c(\phi), \psi \rangle_{\frac{1}{2},\Gamma_E} = \langle \phi, T_c^*(\psi) \rangle_{\frac{1}{2},\Gamma_E}$$

for $\phi, \psi \in H^{1/2}(\Gamma_E)$. Analogously, T_c^* is given by

$$T_c^*(u) = \sum_{n=0}^{\infty} \overline{i\mu_n} \frac{\overline{1 - Z_{0,P}^n}}{\overline{1 + Z_{0,P}^n}} u_n Y_n.$$

4.2. Well-posedness

Now, we consider the problem equipped with CRBC instead of the exact DtN boundary condition on the artificial boundary Γ_E,

$$\begin{aligned} -\Delta u - k^2 u &= f \text{ in } \Omega, \\ \frac{\partial u}{\partial \nu} &= 0 \text{ on } \Gamma_c, \\ \frac{\partial u}{\partial \nu} - T_c(u) &= g \text{ on } \Gamma_E. \end{aligned} \tag{4.4}$$

The corresponding variational problem can be written as finding $u \in H^1(\Omega)$ satisfying

$$A_c(u, \phi) = \langle f, \phi \rangle_{1,\Omega} + \langle g, \phi \rangle_{\frac{1}{2},\Gamma_E} \text{ for all } \phi \in H^1(\Omega), \tag{4.5}$$

where $A_c(\cdot, \cdot)$ is a sesquilinear form defined in $H^1(\Omega) \times H^1(\Omega)$ by

$$A_c(u, \phi) = (\nabla u, \nabla \phi)_\Omega - k^2 (u, \phi)_\Omega - \langle T_c(u), \phi \rangle_{\frac{1}{2},\Gamma_E}.$$

To analyze the well-posedness of the problem (4.5), we will study inf-sup conditions for the sesquilinear form $A_c(\cdot, \cdot)$. If T_c converged to T as operators from $H^{1/2}(\Gamma_E)$ to $H^{-1/2}(\Gamma_E)$ as n_p, $n_e \to \infty$, then the inf-sup conditions of the sesquilinear form $A_c(\cdot, \cdot)$ would be obtained from the inf-sup conditions of $A(\cdot, \cdot)$ by a simple kick-back argument. However this approach is not available

since T_c does not converge to T in norm. Instead we use the convergence of $T_c(\phi)$ to $T(\phi)$ if ϕ is a trace on Γ_E of a radiation solution, which is established in the sequence of lemmas. Let $P_\ell : H^{1/2}(\Gamma_E) \to H^{1/2}(\Gamma_E)$ be the projection onto the finite dimensional subspace spanned by $\{Y_n\}_{n=0}^{\ell}$, i.e.,

$$P_\ell \left(\sum_{n=0}^{\infty} u_n Y_n \right) = \sum_{n=0}^{\ell} u_n Y_n.$$

The first lemma is concerned with the convergence of T_c to T for the propagating component as $n_p \to \infty$.

Lemma 4.2. *Assume that the imaginary parameters a_j and $\tilde{a}_j$ for $0 \le j < n_p$ are the solution of the min-max problem* (2.22)*. Then it holds that*

$$\|(T_c - T)P_N(\phi)\|_{H^{-1/2}(\Gamma_E)} \le Ce^{-Cn_p/\ln(1/\gamma_p)} \|\phi\|_{H^{1/2}(\Gamma_E)}$$

for $\phi \in H^{1/2}(\Gamma_E)$ and for sufficiently large n_p.

Proof. For $\phi = \sum_{n=0}^{\infty} \phi_n Y_n \in H^{1/2}(\Gamma_E)$, by (2.28) and $|i\mu_n|^2 \le C(1+\lambda_n^2)$ we show that

$$\begin{aligned} \|(T_c - T)P_N(\phi)\|^2_{H^{-1/2}(\Gamma_E)} &= \sum_{n=0}^{N} (1+\lambda_n^2)^{-1/2} \left| i\mu_n \frac{2Z^n_{n,p}}{1+Z^n_{0,P}} \phi_n \right|^2 \\ &\le \frac{Ce^{-Cn_p/\ln(1/\gamma_p)}}{(1-e^{-Cn_p/\ln(1/\gamma_p)})^2} \sum_{n=0}^{N} (1+\lambda_n^2)^{1/2} |\phi_n|^2 \\ &\le Ce^{-Cn_p/\ln(1/\gamma_p)} \|\phi\|^2_{H^{1/2}(\Gamma_E)} \end{aligned}$$

for sufficient large $n_p \ge 1$ such that $e^{-Cn_p/\ln(1/\gamma_p)} < 1/2$. □

Under the assumption that ϕ is a trace of a radiating solution on Γ_E, we prove the convergence of T_c for the evanescent component in two lemmas. The first lemma is associated with finite superpositions of relatively slowly decaying evanescent modes.

Lemma 4.3. *Assume that $\phi \in H^{1/2}(\Gamma_E)$ is a trace on Γ_E of a radiating solution $u \in H^1(\Omega)$ to the problem* (4.1)*. Also, assume that the real parameters a_j and $\tilde{a}_j$ for $n_p \le j < n_p + n_e$ are the solution to the the min-max problem* (2.23) *with $\tilde{\mu}_{max} = \tilde{\mu}_M$ for $M > N$. Then it holds that*

$$\|(T_c - T)(P_M - P_N)(\phi)\|_{H^{-1/2}(\Gamma_E)} \le Ce^{-\tilde{\mu}_{N+1}\delta} e^{-Cn_e/\ln(1/\gamma_e)} \|u\|_{H^1(\Omega)}$$

for sufficiently large $n_e \ge 1$ such that $e^{-Cn_e/\ln(1/\gamma_e)} < 1/2$.

Proof. Let u be the radiating solution to the problem (4.1) of the form

$$u(x,y) = \sum_{n=0}^{\infty} A_n e^{i\mu_n(x+\delta)} Y_n(y) \quad \text{for } x > -\delta. \tag{4.6}$$

Then ϕ can be written as

$$\phi = \sum_{n=0}^{\infty} \phi_n Y_n = \sum_{n=0}^{\infty} A_n e^{i\mu_n \delta} Y_n \quad \text{on} \quad H^{1/2}(\Gamma_E). \tag{4.7}$$

As done in Lemma 4.2, we can show that

$$\begin{aligned} \|(T_c - T)(P_M - P_N)(\phi)\|^2_{H^{-1/2}(\Gamma_E)} &= \sum_{n=N+1}^{M} (1+\lambda_n^2)^{-1/2} \left| i\mu_n \frac{2Z^n_{n,p}}{1+Z^n_{0,P}} \phi_n \right|^2 \\ &\le C e^{-Cn_e/\ln(1/\gamma_e)} \sum_{n=N+1}^{M} (1+\lambda_n^2)^{1/2} |\phi_n|^2 \end{aligned}$$

for sufficient large $n_e \ge 1$ such that $e^{-Cn_e/\ln(1/\gamma_e)} < 1/2$. Therefore, by using (4.7) and a trace inequality we obtain that

$$\begin{aligned} \|(T_c - T)(P_M - P_N)(\phi)\|^2_{H^{-1/2}(\Gamma_E)} &\le C e^{-Cn_e/\ln(1/\gamma_e)} \sum_{n=N+1}^{M} (1+\lambda_n^2)^{1/2} e^{-2\tilde{\mu}_n \delta} |A_n|^2 \\ &\le C e^{-2\tilde{\mu}_{N+1}\delta} e^{-Cn_e/\ln(1/\gamma_e)} \|u\|^2_{H^1(\Omega)}, \end{aligned}$$

which completes the proof. □

Finally, we prove the convergence of T_c for remaining evanescent modes.

Lemma 4.4. *Let $\phi \in H^{1/2}(\Gamma_E)$ be a trace on Γ_E of a radiating solution $u \in H^1(\Omega)$ to the problem* (4.1)*. Assume that $M > N$ and $n_e \ge 1$ are determined by the parameter selection procedure for $n_p \ge 1$ so that* (3.1)-(3.3) *hold, and let a_j and $\tilde{a}_j$ be the parameters satisfying the min-max problems* (2.22) *and* (2.23) *with the conditions* (2.10) *and $\tilde{\mu}_{max} = \tilde{\mu}_M$. Then it holds that*

$$\|(T_c - T)(I - P_M)\phi\|_{H^{-1/2}(\Gamma_E)} \le C e^{-\tilde{\mu}_{M+1}\delta} \|u\|_{H^1(\Omega)}. \tag{4.8}$$

Proof. Assume that u and ϕ are written as (4.6) and (4.7). By Lemma 3.3, (4.7) and the fact that $|i\mu_n|^2 \le C(1+\lambda_n^2)$ and $|Z^n_{0,P}| < 1$, we are led to

$$\begin{aligned} \|(T_c - T)(I - P_M)\phi\|^2_{H^{-1/2}(\Gamma_E)} &= \sum_{n=M+1}^{\infty} (1+\lambda_n^2)^{-1/2} \left| i\mu_n \frac{2Z^n_{0,P}}{1+Z^n_{0,P}} \phi_n \right|^2 \\ &\le C \sum_{n=M+1}^{\infty} (1+\lambda_n^2)^{1/2} e^{-2\tilde{\mu}_n \delta} |A_n|^2. \end{aligned}$$

Thus, we can see that by a trace inequality

$$\begin{aligned} \|(T_c - T)(I - P_M)\phi\|^2_{H^{-1/2}(\Gamma_E)} &\le C e^{-2\tilde{\mu}_{M+1}\delta} \sum_{n=M+1}^{\infty} (1+\lambda_n^2)^{1/2} |A_n|^2 \\ &\le C e^{-2\tilde{\mu}_{M+1}\delta} \|u\|^2_{H^1(\Omega)} \end{aligned}$$

which is the required estimate.

□

For the exponential convergence result we denote

$$\mathcal{E}(n_p, n_e, M) = C(e^{-Cn_p/\ln(1/\gamma_p)} + e^{-\tilde{\mu}_{N+1}\delta} e^{-Cn_e/\ln(1/\gamma_e)} + e^{-\tilde{\mu}_{M+1}\delta}).$$

Now combining Lemma 4.2, 4.3 and 4.4, leads to the following convergence result.

Lemma 4.5. *Let $\phi \in H^{1/2}(\Gamma_E)$ be a trace on Γ_E of a radiating solution $u \in H^1(\Omega)$ to the problem (4.1). Assume that $M > N$ and $n_e \geq 1$ are determined by the parameter selection procedure for $n_p \geq 1$ so that (3.1)-(3.3) hold, and let a_j and $\tilde{a}_j$ be the parameters satisfying the min-max problems (2.22) and (2.23) with the conditions (2.10) and $\tilde{\mu}_{max} = \tilde{\mu}_M$ and*

$$e^{-Cn_p/\ln(1/\gamma_p)} < 1/2 \quad \text{and} \quad e^{-Cn_e/\ln(1/\gamma_e)} < 1/2.$$

Then it holds that

$$\|(T_c - T)\phi\|_{H^{-1/2}(\Gamma_E)} \leq \mathcal{E}(n_p, n_e, M)\|u\|_{H^1(\Omega)}. \tag{4.9}$$

The convergence of the adjoint operator holds as well.

Lemma 4.6. *Assume the same conditions as those in Lemma 4.5. Then the same convergence result for the adjoint operators holds.*

$$\|(T_c^* - T^*)\phi\|_{H^{-1/2}(\Gamma_E)} \leq \mathcal{E}(n_p, n_e, M)\|u\|_{H^1(\Omega)}. \tag{4.10}$$

As the last ingredient for the well-posedness analysis, we investigate T_c by breaking into two parts

$$T_c = T_c^1 + T_c^2$$

defined by

$$T_c^1 u = T_c(I - P_N)u = \sum_{n=N+1}^{\infty} i\mu_n \frac{1 - Z_{0,P}^n}{1 + Z_{0,P}^n} u_n,$$

$$T_c^2 u = T_c P_N u = \sum_{n=0}^{N} i\mu_n \frac{1 - Z_{0,P}^n}{1 + Z_{0,P}^n} u_n$$

in the following lemma.

Lemma 4.7. *Assume that n_p is chosen such that $e^{-Cn_p/\ln(1/\gamma_p)} < 1/2$. The real part of the operator T_c^1 is non-positive in $H^{1/2}(\Gamma_E)$ in the sense of*

$$\Re(\langle T_c^1(u), u\rangle_{\frac{1}{2},\Gamma_E}) \leq 0 \tag{4.11}$$

for $u \in H^{1/2}(\Gamma_E)$. Also, it holds that

$$|\langle T_c^2(u), u\rangle_{\frac{1}{2},\Gamma_E}| \leq C\|u\|^2_{L^2(\Gamma_E)} \tag{4.12}$$

for $u \in H^{1/2}(\Gamma_E)$.

Proof. Let $u = \sum_{n=0}^{\infty} u_n Y_n \in H^{1/2}(\Gamma_E)$. Since $|Z_{0,P}^n| < 1$ for all n, we have

$$\Re((1 - Z_{0,P}^n)/(1 + Z_{0,P}^n)) > 0$$

for all n and thereby we can show that

$$\Re(\langle T_c^1(u), u\rangle_{\frac{1}{2},\Gamma_E}) = \sum_{n=N+1}^{\infty} (-\tilde{\mu}_n)\Re\left(\frac{1 - Z_{0,P}^n}{1 + Z_{0,P}^n}\right)|u_n|^2 \leq 0.$$

For establishing (4.12), we invoke that $|Z_{0,P}^n| \leq e^{-Cn_p/\ln(1/\gamma_p)} < 1/2$ for $0 \leq n < N$ to see that

$$\left| i\mu_n \frac{1 - Z_{0,P}^n}{1 + Z_{0,P}^n} \right| < 3k.$$

It then follows that

$$|\langle T_c^2(u), u\rangle_{\frac{1}{2},\Gamma_E}| \leq \sum_{n=0}^{N} \left| i\mu_n \frac{1 - Z_{0,P}^n}{1 + Z_{0,P}^n} \right| |u_n|^2 \leq C\|u\|_{L^2(\Gamma_E)}^2,$$

which completes the proof. □

We are now in a position to prove the well-posedness of the problem (4.5).

Theorem 4.8. *Assume that k^2 is not an eigenvalue of the problem* (2.1)*. Also, assume that $M > N$ and $n_e \geq 1$ are determined by the parameter selection procedure for large n_p and the reflection errors satisfy*

$$e^{-Cn_p/\ln(1/\gamma_p)} < 1/2 \quad \textit{and} \quad e^{-Cn_e/\ln(1/\gamma_e)} < 1/2.$$

Then it holds that

$$\|u\|_{H^1(\Omega)} \leq C \sup_{0 \neq \phi \in H^1(\Omega)} \frac{|A_c(u, \phi)|}{\|\phi\|_{H^1(\Omega)}} \tag{4.13}$$

and

$$\|u\|_{H^1(\Omega)} \leq C \sup_{0 \neq \phi \in H^1(\Omega)} \frac{|A_c(\phi, u)|}{\|\phi\|_{H^1(\Omega)}} \tag{4.14}$$

for all $u \in H^1(\Omega)$. Therefore, the problem (4.5) *has a unique solution $u \in H^1(\Omega)$ such that*

$$\|u\|_{H^1(\Omega)} \leq C(\|f\|_{H^{-1}(\Omega)} + \|g\|_{H^{-1/2}(\Gamma_E)}).$$

Proof. We only prove (4.13), in that (4.14) follows from (4.13) and the fact $A_c(u, \phi) = A_c(\bar{\phi}, \bar{u})$. To prove (4.13), it suffices to show that if $u \in H^1(\Omega)$ satisfies

$$A_c(u, \phi) = F(\phi) \text{ for } \phi \in H^1(\Omega) \tag{4.15}$$

for some $F \in H^{-1}(\Omega)$, then it holds that

$$\|u\|_{H^1(\Omega)} \leq C\|F\|_{H^{-1}(\Omega)}. \tag{4.16}$$

We start by examining the real part of $A_c(u,u)$,

$$\Re(A_c(u,u)) = \|u\|^2_{H^1(\Omega)} - (k^2+1)\|u\|^2_{L^2(\Omega)} - \Re(\langle T_c^1(u), u\rangle_{\frac{1}{2},\Gamma_E}) - \Re(\langle T_c^2(u), u\rangle_{\frac{1}{2},\Gamma_E}).$$

From (4.11) and (4.12) in Lemma 4.7, it follows that

$$\|u\|^2_{H^1(\Omega)} \le |A_c(u,u)| + (k^2+1)\|u\|^2_{L^2(\Omega)} + C\|u\|^2_{L^2(\Gamma_E)}. \tag{4.17}$$

From now on, we shall estimate three terms in the right hand side of (4.17). The first term is easily estimated by

$$|A_c(u,u)| = |F(u)| \le \|F\|_{H^{-1}(\Omega)}\|u\|_{H^1(\Omega)}. \tag{4.18}$$

For the second and third terms, we consider a solution $w \in H^1(\Omega)$ to the adjoint problem

$$A(\phi, w) = (\phi, u)_\Omega + (\phi, u)_{\Gamma_E} \text{ for all } \phi \in H^1(\Omega), \tag{4.19}$$

satisfying

$$\|w\|_{H^1(\Omega)} \le C(\|u\|_{L^2(\Omega)} + \|u\|_{L^2(\Gamma_E)}). \tag{4.20}$$

Here we note that $\bar{w}$ is a radiating solution. By taking $\phi = u$ in (4.19), using Lemma 4.6, (4.20) and a trace inequality we show that

$$\begin{aligned}
\|u\|^2_{L^2(\Omega)} + \|u\|^2_{L^2(\Gamma_E)} &= A(u,w) = A_c(u,w) + \langle (T_c - T)u, w\rangle_{\Gamma_E} \\
&= F(w) + \langle u, (T_c^* - T^*)w\rangle_{\Gamma_E} \\
&\le \|F\|_{H^{-1}(\Omega)}\|w\|_{H^1(\Omega)} + \mathcal{E}(n_p, n_e, M)\|u\|_{H^1(\Omega)}\|w\|_{H^1(\Omega)} \\
&\le \left(\|F\|_{H^{-1}(\Omega)} + \mathcal{E}(n_p, n_e, M)\|u\|_{H^1(\Omega)}\right)(\|u\|_{L^2(\Omega)} + \|u\|_{L^2(\Gamma_E)}).
\end{aligned}$$

As a consequence, it can be obtained that

$$(k^2+1)\|u\|^2_{L^2(\Omega)} + C\|u\|^2_{L^2(\Gamma_E)} \le C\left(\|F\|^2_{H^{-1}(\Omega)} + \mathcal{E}(n_p, n_e, M)^2\|u\|^2_{H^1(\Omega)}\right). \tag{4.21}$$

Now, we use (4.18) and (4.21) in (4.17) and apply the arithmetic-geometric mean inequality to (4.18) with $\gamma > 0$ to have

$$\|u\|^2_{H^1(\Omega)} \le \left(\frac{\gamma}{2} + \mathcal{E}(n_p, n_e, M)^2\right)\|u\|^2_{H^1(\Omega)} + \left(\frac{1}{2\gamma} + C\right)\|F\|^2_{H^{-1}(\Omega)}.$$

Finally, by choosing small γ and large n_p,n_e and M such that $\gamma/2 + \mathcal{E}(n_p, n_e, M)^2 < 1/2$, (4.16) then follows, which completes the proof.

□

4.3. Convergence

Now, we are ready to show the convergence of approximate solutions satisfying CRBCs as n_p and n_e tend towards infinity.

Theorem 4.9. *Assume the same conditions as those in Theorem 4.8. Let u^{ex} and u be the solutions to the problem* (4.2) *and* (4.5)*, respectively. Then for sufficiently large n_p, n_e and M, we have*

$$\|u - u^{ex}\|_{H^1(\Omega)} \leq \mathcal{E}(n_p, n_e, M)\|u^{ex}\|_{H^1(\Omega)}.$$

Proof. Let $e = u - u^{ex}$ be the error function. Then e solves the problem

$$A_c(e, \phi) = \langle (T_c - T)u^{ex}, \phi \rangle_{\frac{1}{2}, \Gamma_E} \text{ for } \phi \in H^1(\Omega). \tag{4.22}$$

From Theorem 4.8 and Lemma 4.5, it can be derived that

$$\|e\|_{H^1(\Omega)} \leq C\|(T_c - T)u^{ex}\|_{H^{-1/2}(\Gamma_E)} \leq \mathcal{E}(n_p, n_e, M)\|u^{ex}\|_{H^1(\Omega)},$$

which is the desired convergence result. □

Remark 4.10. *We provide the well-posedness and convergence analysis only for* CRBC *with optimally chosen parameters for a concise presentation. However it still holds when some of parameters are equal to $-i\mu_n$ of important modes. The analysis for this case requires only a minor modification with regard to the boundedness* (3.9)*.*

1. *For small tolerance $\varepsilon > 0$, let $M > N$ be an integer such that $e^{-\tilde{\mu}_{M+1}\delta} < \varepsilon$. We consider the* CRBC *giving the exact radiation condition to the first $M+1$ modes, that is, all $-i\mu_n$ for $0 \leq n \leq M$ are used for* CRBC *parameters. In this case* (3.9) *holds. Indeed, this case yields that $\tilde{\mu}_{ang}$ can be taken to be $6Na_{av}/\pi$ with a_{av} being the arithmetic mean of the axial frequency of propagating modes, and it is a constant independent of M. Thus, if $\tilde{\mu}_{M+1} > \tilde{\mu}_{ang}$, then $|1 + Z^n_{0,P}| > 1/2$ for all $n > M$. In case that $\tilde{\mu}_{M+1} \leq \tilde{\mu}_{ang}$, since there are only finite number of $\tilde{\mu}_n$ such that $\tilde{\mu}_{M+1} < \tilde{\mu}_n < \tilde{\mu}_{ang}$, we can find a constant C_M such that $|1 + Z^n_{0,P}| > C_M > 0$ for $\tilde{\mu}_{M+1} < \tilde{\mu}_n < \tilde{\mu}_{ang}$. Therefore, it can be concluded that $|1 + Z^n_{0,P}| > \min\{C_M, 1/2\}$, where the minimum is taken over M such that $\tilde{\mu}_{M+1} \leq \tilde{\mu}_{ang}$. The convergence result of this case can be written as*

$$\|u - u^{ex}\|_{H^1(\Omega)} \leq Ce^{-\tilde{\mu}_{M+1}\delta}\|u^{ex}\|_{H^1(\Omega)}.$$

2. *We consider the* CRBC *eliminating the reflection of near-cutoff modes and minimizing those of other modes by using optimally chosen parameters, that is, $-i\mu_n$ of near-cutoff modes is used as a parameter and other parameters are determined by the formulas* (2.26)*. We can have* (3.9) *as the counterparts of* (3.11) *and* (3.13) *follow the analogous asymptotic behaviors. The convergence result of this case can be written as*

$$\|u - u^{ex}\|_{H^1(\Omega)} \leq C(e^{Cn'_p/\ln(1/\gamma'_p)} + e^{Cn'_e/\ln(1/\gamma'_e)} + e^{-\tilde{\mu}_{M+1}\delta})\|u^{ex}\|_{H^1(\Omega)},$$

where

$$(n'_p,\ \gamma'_p,\ n'_e,\ \gamma'_e) = \begin{cases} (n_p - 1,\ k/\mu_{N-2},\ n_e,\ \gamma_e) \text{ or} \\ (n_p,\ \gamma_p,\ n_e - 1,\ \tilde{\mu}_{max}/\tilde{\mu}_{N+2}) \end{cases}$$

depending on whether the near-cutoff mode is a propagating mode or an evanescent mode.

5. Practical CRBCs for discretization methods

In order to apply a discretization technique such as the finite element method to the problem (2.17) with CRBC imposed on Γ_E, we need to rephrase the boundary condition in a tractable form. To this end, we introduce a practical CRBC involving auxiliary variables defined only on the boundary Γ_E. Let $\boldsymbol{e}_j$ be the j-th standard basis vector in $\mathbb{C}^{P+2}$ for $j = 0, 1, \ldots, P+1$, whose nonzero entry is one at the j-th component, and let us define a space for auxiliary variables,

$$\boldsymbol{V}_{\Gamma_E} = \{(\phi_0, \ldots, \phi_{P+1}) \in (L^2(\Gamma_E))^{P+2} \ : \ \phi_j + \phi_{j+1} \in H^1(\Gamma_E) \text{ for } j = 0, \ldots, P\}.$$

After eliminating the normal derivatives of auxiliary variables in the recursive relations (2.7) as done in [17], we can see that if u satisfies CRBC on Γ_E, then there exists $\Phi = (\phi_0, \ldots, \phi_{P+1})^t \in \boldsymbol{V}_{\Gamma_E}$ satisfying

$$\begin{aligned} \frac{\partial u}{\partial \nu}\boldsymbol{e}_0 &= -L\frac{\partial^2}{\partial \nu^2}\Phi - M\Phi \\ &= L\nabla_y^2\Phi + (k^2 L - M)\Phi \text{ on } \Gamma_E \end{aligned} \tag{5.1}$$

and $\frac{\partial}{\partial \nu}\Phi = 0$ on $\partial\Gamma_E$ with $\phi_0 = u$ on Γ_E, where L and M are $(P+2)\times(P+2)$ symmetric and tridiagonal matrices whose nonzero entries are

$$L_{j,j-1} = \frac{1}{a_{j-1} + \tilde{a}_{j-1}}, \quad L_{j,j} = \frac{1}{a_{j-1} + \tilde{a}_{j-1}} + \frac{1}{a_j + \tilde{a}_j}, \quad L_{j,j+1} = \frac{1}{a_j + \tilde{a}_j},$$
$$M_{j,j-1} = \frac{-a_{j-1}^2}{a_{j-1} + \tilde{a}_{j-1}}, \quad M_{j,j} = \frac{a_{j-1}\tilde{a}_{j-1}}{a_{j-1} + \tilde{a}_{j-1}} + \frac{a_j \tilde{a}_j}{a_j + \tilde{a}_j}, \quad M_{j,j+1} = \frac{-\tilde{a}_j^2}{a_j + \tilde{a}_j}$$

for $j = 0, \ldots, P+1$. Here we use the convention that the terms with indices outside the bounds of the arrays a_j and $\tilde{a}_j$ are ignored, for instance,

$$L_{0,0} = \frac{1}{a_0 + \tilde{a}_0} \quad \text{and} \quad L_{P+1,P+1} = \frac{1}{a_P + \tilde{a}_P}.$$

Therefore one can have the problem suitable for application of discretization methods, to find $u \in H^1(\Omega)$ and $\Phi \in \boldsymbol{V}_{\Gamma_E}$ satisfying

$$\begin{aligned} -\Delta u - k^2 u &= f \text{ in } \Omega, \\ \frac{\partial u}{\partial \nu} &= 0 \text{ on } \Gamma_c, \\ \frac{\partial u}{\partial \nu}\boldsymbol{e}_0 &= L\nabla_y^2\Phi + (k^2 L - M)\Phi \text{ on } \Gamma_E \end{aligned} \tag{5.2}$$

with $u = \phi_0$ and $\partial\Phi/\partial\nu = 0$ on $\partial\Gamma_E$. By introducing the solution space

$$\boldsymbol{V} = \{(u, \Phi) \in H^1(\Omega) \times \boldsymbol{V}_{\Gamma_E} \ : \ u = \phi_0 \text{ on } \Gamma_E \text{ for } \Phi = (\phi_0, \ldots, \phi_{P+1})\}$$

equipped with the weighted norm depending on the matrix L

$$\|(u, \Phi)\|_{\boldsymbol{V}}^2 := \|u\|_{H^1(\Omega)}^2 + \|\Phi\|_{\boldsymbol{V}_{\Gamma_E}}^2,$$

where

$$\|\Phi\|^2_{\boldsymbol{V}_{\Gamma_E}} := \sum_{j=0}^{P} \frac{1}{|a_j + \tilde{a}_j|} \|\phi_j + \phi_{j+1}\|^2_{H^1(\Gamma_E)} \tag{5.3}$$

we can rewrite the problem (5.2) as a variational problem to seek for $(u, \Phi) \in \boldsymbol{V}$ satisfying

$$(\nabla u, \nabla v)_\Omega - k^2(u, v)_\Omega + (L\nabla_y \Phi, \nabla_y \Psi)_{\Gamma_E} + ((-k^2 L + M)\Phi, \Psi)_{\Gamma_E} = \langle f, v\rangle_{1,\Omega} \tag{5.4}$$

for all $(v, \Psi) \in \boldsymbol{V}$.

The equivalence of two problems (2.17) and (5.2) is established in the following theorem.

Theorem 5.1. *If $(u, \Phi) \in \boldsymbol{V}$ is a solution to the problem (5.2), then u is a solution to the problem (2.17). Conversely, if $u \in H^1(\Omega)$ is a solution to the problem (2.17), then there exists $\Phi \in \boldsymbol{V}_{\Gamma_E}$ such that (u, Φ) is a solution in $\boldsymbol{V}$ to the problem (5.2). In addition, the auxiliary variables Φ to the problem (5.2) satisfy*

$$\sqrt{\hat{\mu}_{min}} \|\Phi\|_{\boldsymbol{V}_{\Gamma_E}} \le C\sqrt{P+1} \|f\|_{H^{-1}(\Omega)}, \tag{5.5}$$

where $\hat{\mu}_{min} := \min\{|\mu_n| \ : \ \mu_n \neq 0\}$ is the smallest non-zero axial frequency.

To prove Theorem 5.1, we first show the n-th mode of the solution to the problem (5.2) satisfies the same DtN boundary condition as (2.19). To do this, we examine the auxiliary variables solving the problem on Γ_E

$$\begin{aligned} -L\nabla_y^2 \Phi + (-k^2 L + M)\Phi &= E_j \boldsymbol{e}_j \text{ in } \Gamma_E, \\ \frac{\partial \Phi}{\partial \nu} &= 0 \text{ on } \partial\Gamma_E \end{aligned} \tag{5.6}$$

for $E_j \in H^{-1}(\Gamma_E)$ with $0 \le j \le P+1$. Although it is enough to study the problem (5.6) with $j = 0$, we examine the general problem for the unique solvability of the problem (5.6).

It is clear that the n-th Fourier coefficients Φ^n of Φ satisfy the equation

$$(-\mu_n^2 L + M)\Phi^n = E_j^n \boldsymbol{e}_j, \tag{5.7}$$

where E_j^n is the n-th Fourier coefficient of E_j. We shall establish the invertibility of the system matrix of (5.7) for $n \neq N$ and derive the solution formulas in the following lemma.

Lemma 5.2. *Let $n \neq N$, i.e., $\mu_n \neq 0$. Then $-\mu_n^2 L + M$ is invertible and a solution $\Phi^n \in \mathbb{C}^{P+2}$ to the linear system (5.7) is given by the formula $\phi_\ell^n = s_{\ell,j}^n E_j^n$, where*

$$s_{\ell,j}^n = \begin{cases} \dfrac{(1 + Z_{0,\ell-1}^n) R_{\ell,j-1}^n (1 + Z_{j,P}^n)}{-2i\mu_n (1 - Z_{0,P}^n)} & \text{if } \ell \le j, \\ \dfrac{(1 + Z_{0,j-1}^n) Q_{j,\ell-1}^n (1 + Z_{\ell,P}^n)}{-2i\mu_n (1 - Z_{0,P}^n)} & \text{if } \ell \ge j. \end{cases} \tag{5.8}$$

Proof. Once we show that the problem (5.7) for each $j = 0, 1, \ldots, P+1$ has a solution, the invertibility of the matrix $-\mu_n^2 L + M$ immediately follows.

We first assume that $\tilde{a}_j \neq -i\mu_n$ for $j = 0, 1, \ldots, P$. In case that $0 < j < P+1$, we will find the solution Φ^n of the form

$$\phi_\ell^n = \begin{cases} Q_{0,\ell-1}^n \widetilde{A}_n + (R_{0,\ell-1}^n)^{-1} \widetilde{B}_n & \text{for } \ell = 0, 1, \ldots, j, \\ Q_{j,\ell-1}^n \widetilde{C}_n + (R_{j,\ell-1}^n)^{-1} \widetilde{D}_n & \text{for } \ell = j, j+1, \ldots, P+1. \end{cases} \tag{5.9}$$

When $j = 0$ or $P+1$, we will find solution ϕ_ℓ^n defined by the upper formula with $\ell = 0, 1, \ldots, P+1$. Here we will verify the formulas for $0 < j < P+1$, as the other cases can be treated in the same way with only small modifications.

By the definition of $Q_{j,m}^n$ and $R_{j,m}^n$ one can easily show that the three term recursions

$$(-\mu_n^2 L_{\ell,\ell-1} + M_{\ell,\ell-1})\phi_{\ell-1}^n + (-\mu_n^2 L_{\ell,\ell} + M_{\ell,\ell})\phi_\ell^n + (-\mu_n^2 L_{\ell,\ell+1} + M_{\ell,\ell+1})\phi_{\ell+1}^n = 0$$

hold for $\ell \neq 0, j, P+1$. Thus, the four unknowns $\widetilde{A}_n$, $\widetilde{B}_n$, $\widetilde{C}_n$ and $\widetilde{D}_n$ are to be determined by

$$-i\mu_n(\widetilde{A}_n - \widetilde{B}_n) = 0 \tag{5.10}$$

from the 0-th equation,

$$Q_{0,j-1}^n \widetilde{A}_n + (R_{0,j-1}^n)^{-1} \widetilde{B}_n = \widetilde{C}_n + \widetilde{D}_n \tag{5.11}$$

from the definition of ϕ_ℓ^n with $\ell = j$,

$$(Q_{0,j-1}^n \widetilde{A}_n - (R_{0,j-1}^n)^{-1} \widetilde{B}_n) - (\widetilde{C}_n - \widetilde{D}_n) = \frac{1}{i\mu_n} E_j^n \tag{5.12}$$

from the j-th equation and

$$Q_{j,P}^n \widetilde{C}_n - (R_{j,P}^n)^{-1} \widetilde{D}_n = 0 \tag{5.13}$$

from the $(P+1)$-th equation. Solving the equations (5.10)-(5.13) leads to

$$\widetilde{A}_n = \widetilde{B}_n = \frac{(1 + Z_{j,P}^n) R_{0,j-1}^n}{-2i\mu_n(1 - Z_{0,P}^n)} E_j^n,$$
$$\widetilde{C}_n = \frac{(1 + Z_{0,j-1}^n)}{-2i\mu_n(1 - Z_{0,P}^n)} E_j^n, \qquad \widetilde{D}_n = \frac{(1 + Z_{0,j-1}^n) Z_{j,P}^n}{-2i\mu_n(1 - Z_{0,P}^n)} E_j^n,$$

and hence the formula (5.8) is obtained. When there exists an index J such that $\tilde{a}_J + i\mu_n = 0$, it is straightforward to show directly that $\Phi^n = E_j^n \sum_{\ell=0}^P s_{\ell,j}^n \boldsymbol{e}_\ell$ still solves the problem (5.7), which completes the proof.

□

The next lemma is required to analyze cutoff modes associated with $\mu_N = 0$.

Lemma 5.3. *$\boldsymbol{e}_0$ is not in the range of M, that is, $M\Phi^N = \boldsymbol{e}_0$ does not have a solution.*

Proof. We conduct backward Gaussian elimination to reduce the tridiagonal matrix M to the lower triangular matrix. It turns out that the reduced lower triangular matrix has zero at the $(0,0)$-entry and hence the problem $M\Phi^N = \boldsymbol{e}_0$ does not admit a solution, which completes the proof. □

For the estimate (5.5) of the auxiliary variables solving the problem (5.2) we need following lemmas.

Lemma 5.4. *Assume that a_j and $\tilde{a}_j$ are the solution parameters defined by* (2.26) *for the min-max problems* (2.22) *and* (2.23)*, and n_p and n_e grow at the rate of* (3.1)-(3.3)*. For each m let $C_m = \max\{|\tilde{a}_j/a_j|\}_{j=0}^{m-1} > 1$, where $m = n_p$ or $m = n_e$. Then $C_m \to 1$ as $m \to \infty$.*

Proof. It suffices to work with s_j of (2.25). Let $\Delta = K(\tilde{\gamma})/m$ and $u_j = (1 - (2j+1)/2m)K(\tilde{\gamma})$. Here we observe that $\Delta \to 0$ as $m \to \infty$. Indeed, in case that $\gamma = \gamma_p$, $m = n_p$ the convergence of Δ is an obvious result from the fact that $K(\tilde{\gamma}_p)$ is constant. On the other hand, if $\gamma = \gamma_e$, $m = n_e$, then we use (3.6), (3.1)-(3.3) to show that

$$\frac{K(\tilde{\gamma})}{m} \asymp \frac{\ln(1/\gamma_e)}{\ln(1/\varepsilon)\cdot \ln\ln(1/\varepsilon)} \asymp \frac{\ln\ln(1/\varepsilon)}{\ln(1/\varepsilon)\cdot \ln\ln(1/\varepsilon)} \to 0$$

as $n_e \to \infty$.

Now by the mean value theorem, we have

$$\frac{s_{j+1}}{s_j} = \frac{\mathrm{dn}(u_j - \Delta, \tilde{\gamma})}{\mathrm{dn}(u_j, \tilde{\gamma})} = 1 + \frac{\tilde{\gamma}\mathrm{sn}(u_j^*, \tilde{\gamma})\mathrm{cn}(u_j^*, \tilde{\gamma})}{\mathrm{dn}(u_j, \tilde{\gamma})}\Delta$$

for some u_j^* between u_j and u_{j+1}. By using $\mathrm{cn}(u_j^*, \tilde{\gamma}) \le \mathrm{dn}(u_j^*, \tilde{\gamma}) \le \mathrm{dn}(u_{j+1}, \tilde{\gamma})$ and $0 < \tilde{\gamma}\mathrm{sn}(u_j^*, \tilde{\gamma}) < 1$, we have

$$\frac{s_{j+1}}{s_j} < 1 + \Delta\frac{s_{j+1}}{s_j}$$

or equivalently, $(1-\Delta)s_{j+1}/s_j < 1$. Since that $\Delta \to 0$ as $m \to \infty$, we see that $C_m < (1-\Delta)^{-1}$ for sufficiently large m and hence $C_m \to 1$ as $m \to \infty$. □

Lemma 5.5. *Under the same assumptions as those in Lemma 5.4, it holds that*

$$|Q_{0,\ell}^n| < 1 \tag{5.14}$$

for $n \neq N$ and for $0 \le \ell \le P$. In addition, there exists a positive constant C independent of n_p, n_e and CRBC *parameters such that*

$$\frac{|1 + Q_{j,j}^n|}{\sqrt{|a_j + \tilde{a}_j|}} \le \frac{C}{\sqrt{|\mu_n|}}, \qquad \frac{|1 + R_{j,j}^n|}{\sqrt{|a_j + \tilde{a}_j|}} \le \frac{C}{\sqrt{|\mu_n|}}. \tag{5.15}$$

Proof. The condition $|a_j| < |\tilde{a}_j|$ in (2.10) leads to $|a_j + i\mu_n|/|\tilde{a}_j - i\mu_n| < 1$ and it then follows that $|Q_{0,\ell}^n| < 1$.

For (5.15) we first observe $\sqrt{|\tilde{a}_j\mu_n|}/|\tilde{a}_j - i\mu_n| \le 1$ by the arithmetic-geometric mean inequality. Utilizing the inequality and $|a_j| < |\tilde{a}_j|$ leads to

$$\begin{aligned}\frac{|1+Q^n_{j,j}|}{\sqrt{|a_j+\tilde{a}_j|}} &= \frac{1}{\sqrt{|a_j+\tilde{a}_j|}}\frac{|a_j+\tilde{a}_j|}{|\tilde{a}_j - i\mu_n|} \\ &= \frac{\sqrt{|\tilde{a}_j\mu_n|}}{|\tilde{a}_j - i\mu_n|}\frac{\sqrt{|a_j+\tilde{a}_j|}}{\sqrt{|\tilde{a}_j|}}\frac{1}{\sqrt{|\mu_n|}} \le \frac{C}{\sqrt{|\mu_n|}},\end{aligned} \tag{5.16}$$

which proves the first inequality of (5.15). The second inequality of (5.15) can be established in the same way but with $|\tilde{a}_j| \le C|a_j|$ from Lemma 5.4 instead of $|a_j| < |\tilde{a}_j|$. Here the constant C is taken to be the maximum of C_m, which is available since C_m converges to 1^+.

□

Now we are ready to prove Theorem 5.1.

Proof of Theorem 5.1. Assume that $(u, \Phi) \in \boldsymbol{V}$ is a solution to the problem (5.2). Lemma 5.3 asserts that there is no contribution of cutoff modes to the Neumann data of u, i.e.,

$$\frac{\partial u_N}{\partial \nu} = 0 \text{ on } \Gamma_E.$$

For non-cutoff modes, the solution formula (5.8) with $\ell = j = 0$ and $u_n = \phi^n_0$ shows that

$$\frac{\partial u_n}{\partial \nu} = \frac{-1}{s^n_{0,0}}u_n = i\mu_n\frac{1-Z^n_{0,P}}{1+Z^n_{0,P}}u_n. \tag{5.17}$$

As a consequence, all modes of the solution u satisfy the condition (2.19) and hence u solves the problem (2.17).

Conversely, assume that $u \in H^1(\Omega)$ is a solution to the problem (2.17) and consider the problem (5.7) with $E^n_0 = -\partial u_n/\partial\nu$. When $n \ne N$, the solution formula (5.8) shows that

$$\phi^n_\ell = \frac{Q^n_{0,\ell-1}(1+Z^n_{\ell,P})}{i\mu_n(1-Z^n_{0,P})}\frac{\partial u_n}{\partial \nu} \quad \text{for } \ell = 0, 1, \dots, P+1. \tag{5.18}$$

By comparing (5.18) of $\ell = 0$ with (5.17) we can show that $\phi^n_0 = u_n$. Furthermore, eliminating $\partial u_n/\partial\nu$ from (5.18) yields that

$$\phi^n_\ell = \frac{Q^n_{0,\ell-1}(1-Z^n_{\ell,P})}{(1+Z^n_{0,P})}\phi^n_0 \quad \text{for } \ell = 0, 1, \dots, P+1.$$

Now, for the estimate (5.5) of Φ, we use Lemma 3.3 and (5.14) of Lemma 5.5 to obtain, for $n \ne N$,

$$\begin{aligned}|\phi^n_\ell + \phi^n_{\ell+1}| &= \left|\frac{Q^n_{0,\ell-1}(1+Q^n_{\ell,\ell}) + Q^n_{0,\ell}Z^n_{\ell+1,P}(1+R^n_{\ell,\ell})}{1+Z^n_{0,P}}\right||\phi^n_0| \\ &\le C(|1+Q^n_{\ell,\ell}| + |1+R^n_{\ell,\ell}|)|\phi^n_0|\end{aligned} \tag{5.19}$$

for $\ell = 0, \ldots, P$ and hence by (5.15) of Lemma 5.5

$$\frac{1}{\sqrt{|a_\ell + \tilde{a}_\ell|}} |\phi_\ell^n + \phi_{\ell+1}^n| \le \frac{C}{\sqrt{|\mu_n|}} |\phi_0^n|. \tag{5.20}$$

For $n = N$, there exists a unique solution to the problem $M\Phi_N = 0$ with $\phi_0^N = u_N$, that is,

$$\phi_\ell^N = \left(\prod_{j=0}^{\ell-1} \frac{a_j}{\tilde{a}_j} \right) \phi_0^N.$$

Therefore, we have

$$|\phi_\ell^N + \phi_{\ell+1}^N| = \left(\prod_{j=0}^{\ell-1} \left| \frac{a_j}{\tilde{a}_j} \right| \right) \frac{|a_\ell + \tilde{a}_\ell|}{|\tilde{a}_\ell|} |\phi_0^N|$$

for $\ell = 0, \ldots, P$ and so by using the inequalities

$$\frac{\sqrt{|a_\ell + \tilde{a}_\ell|}}{|\tilde{a}_\ell|} \le \frac{C}{\sqrt{\hat{\mu}_{min}}} \quad \text{and} \quad \prod_{j=0}^{\ell-1} \left| \frac{a_j}{\tilde{a}_j} \right| \le 1,$$

we are led to

$$\frac{1}{\sqrt{|a_\ell + \tilde{a}_\ell|}} |\phi_\ell^N + \phi_{\ell+1}^N| = \frac{\sqrt{|a_\ell + \tilde{a}_\ell|}}{|\tilde{a}_\ell|} \left(\prod_{j=0}^{\ell-1} \left| \frac{a_j}{\tilde{a}_j} \right| \right) |\phi_0^N| \le \frac{C}{\sqrt{\hat{\mu}_{min}}} |\phi_0^N|. \tag{5.21}$$

Noting that

$$\frac{(1+\lambda_n^2)^{1/2}}{|\mu_n|} \le \frac{C}{\hat{\mu}_{min}}, \tag{5.22}$$

we combine (5.20) and (5.21) to obtain

$$(\lambda_n^2 + 1) \|\Phi^n\|_{\mathcal{L}}^2 \le C \frac{(P+1)}{\hat{\mu}_{min}} (\lambda_n^2 + 1)^{1/2} |\phi_0^n|^2 \tag{5.23}$$

for $n \ge 0$, which results in the desired inequality (5.5)

$$\sqrt{\hat{\mu}_{min}} \|\Phi\|_{\boldsymbol{V}_{\Gamma_E}} \le C\sqrt{P+1} \|\phi_0\|_{H^{1/2}(\Gamma_E)} \le C\sqrt{P+1} \|f\|_{H^{-1}(\Omega)}$$

and the proof is completed.

□

Remark 5.6. *Since $|Q_{0,P}^n| \approx 1$ for large n in (5.19), the dependence of the stability constant in (5.5) on P may arise for such large n. However for relatively small n (such as all propagating modes and relatively slowly decaying evanescent modes) such that*

$$|Q_{j,j}^n| < c < 1 \text{ for all } 0 \le j \le P \tag{5.24}$$

n	$k=10$		$k=16$	
	μ_n	$\lvert e^{i\mu_n\delta}\rvert$	μ_n	$\lvert e^{i\mu_n\delta}\rvert$
0	10	1	16	1
1	9.4937	1	15.6885	1
2	7.7795	1	14.7147	1
3	3.3426	1	12.9296	1
4	$7.6101i$	3.2564e-02	9.9038	1
5	$12.1136i$	4.2914e-03	3.0430	1
6	$15.9783i$	7.5391e-04	$9.9652i$	1.1284e-02
7	$19.5860i$	1.4868e-04	$15.0867i$	1.1260e-03
8	$23.0576i$	3.1173e-05	$19.3818i$	1.6299e-04
9	$26.4469i$	6.7829e-06	$23.3118i$	2.7805e-05
10	$29.7819i$	1.5123e-06	$27.0363i$	5.2027e-06
11	$33.0790i$	3.4299e-07	$30.6304i$	1.0323e-06
12	$36.3486i$	7.8758e-08	$34.1354i$	2.1322e-07
13	$39.5975i$	1.8254e-08	$37.5761i$	4.5332e-08
14	$42.8304i$	4.2613e-09	$40.9688i$	9.8483e-09
$\vdots$			$\vdots$	$\vdots$

Table 3: Magnitude of $e^{i\mu_n\delta}$.

with some positive constant $c<1$ and hence $|Q^n_{0,\ell-1}|<c^\ell$, we see that (5.20) *is bounded by a geometric sequence*

$$\frac{1}{\sqrt{|a_\ell+\tilde{a}_\ell|}}|\phi_\ell+\phi_{\ell+1}|<c^\ell\frac{C}{\sqrt{|\mu_n|}}|\phi_0^n|$$

It implies that the norm estimate (5.23) *is derived without the dependence on P. Thus if the solution consists of modes with small n satisfying* (5.24)*, then the dependence of the stability constant in* (5.5) *on P is not involved.*

6. Numerical examples

This section is devoted to presenting numerical experiments validating the convergence theory for CRBC approximate solutions. In the first example, we show that CRBC can provide the exact radiation condition for important modes. To this end, we choose a semi-infinite straight waveguide with a rectangular obstacle for $\Omega_\infty=((0,\infty)\times(0,1))\setminus\bar{\Omega}_1$ with $\Omega_1=(0.01,0.5)\times(0.01,0.99)$. The unbounded domain is truncated at $x=1$ and the computation domain Ω is set to be $\Omega=(0,1)^2\setminus\bar{\Omega}_1$. For $k=10$ and 16 the wave source f is given in a way that the exact solution u^{ex} is defined by

$$u^{ex}(x,y)=\chi(x)\sum_{n=0}^{5L-1}\frac{1}{L}e^{i\mu_n(x-0.55)}\cos(n\pi y),$$

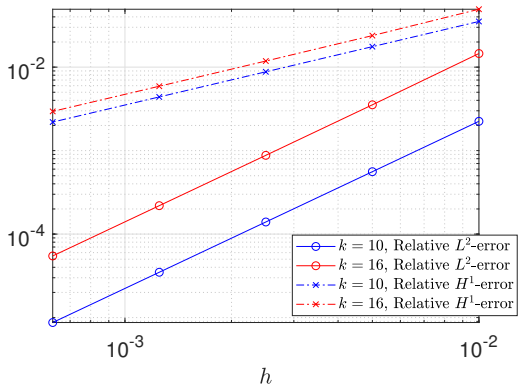


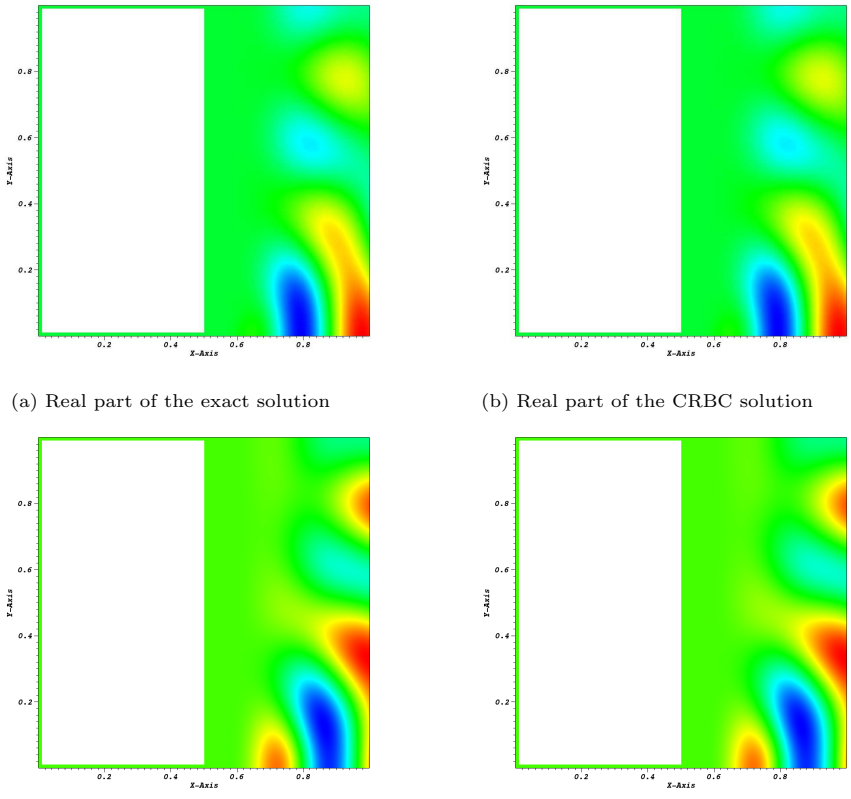
Figure 4: Relative errors of finite element approximations for $k = 10$ and $k = 16$.

where $L = \lfloor k/\pi \rfloor$ and χ is a smooth cutoff function of x such that $\|\chi\|_\infty = 1$, $\chi(x) = 0$ for $x < 0.55$ and $\chi(x) = 1$ for $x > 0.9$. When $k = 10$, we have $L = 3$ and there are 4 propagating modes and 11 evanescent modes in the solution and in case of $k = 16$, $L = 5$ and the solution is a superposition of 6 propagating modes and 19 evanescent modes. The magnitude of $e^{i\mu_n \delta}$ for each mode is given in Table 3 with the separation $\delta = 0.45$ between the artificial boundary and the wave source of evanescent modes. We determine $n_e = 4$ and choose parameters a_j and $\tilde{a}_j$ to be the values of $-i\mu_n$ in the gray cells in the table since the reflection errors of remaining evanescent modes are bounded by

$$|e^{i\mu_{12}\delta}| \approx 7.87581 \times 10^{-8} \text{ for } k = 10,$$
$$|e^{i\mu_{14}\delta}| \approx 9.84837 \times 10^{-9} \text{ for } k = 16,$$

which are small enough for finite element approximations. The CRBC with $(n_p, n_e) = (2, 4)$ for $k = 10$ and $(n_p, n_e) = (3, 4)$ for $k = 16$, whose imaginary parameters are equal to the axial frequencies of all propagating modes and real parameters coincide with the decay rates of the first 8 evanescent modes, serves as the exact radiation condition for those modes. We compute bilinear finite element approximate solutions with mesh $h = 1/100$, $1/200$, $1/400$, $1/800$ and $1/1600$ by using the finite element library deal.II [3]. As seen in Figure 4, finite element approximations converge at the quasi-optimal rate since reflection errors are less than 10^{-7}. The snapshots of real and imaginary parts of the exact solution and the approximate solution satisfying CRBC obtained by the finite element method with $h = 1/400$ are presented in Figure 5, showing that the approximate solution is not visually distinguishable from the exact one.

In many applications, we do not need to choose many parameters to make CRBC an exact radiation condition for all important modes but instead it is enough to reduce reflection errors below finite element errors by using CRBC with a smaller number of optimal parameters. To demonstrate the effect of optimally chosen parameters, we consider a problem for $k = 10\pi$ posed on the



(a) Real part of the exact solution

(b) Real part of the CRBC solution

(c) Imaginary part of the exact solution

(d) Imaginary part of the CRBC solution

Figure 5: Snapshots of the real and imaginary parts of the exact solution and the finite element solution satisfying CRBC for $k = 16$ and $h = 1/400$.

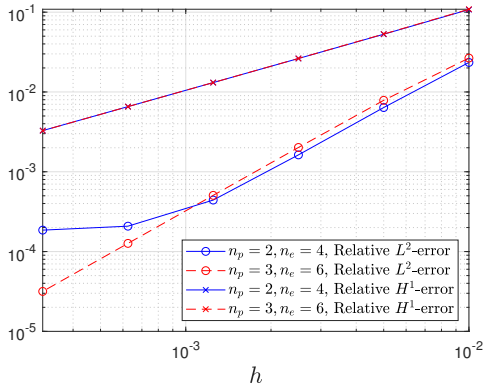


Figure 6: Relative errors of finite element approximations satisfying CRBC for $h = 1/100, 1/200, 1/400, 1/800, 1/1600$ and $1/3200$ when $k = 10\pi$.

computational domain $\Omega = (0, 0.05) \times (0, 1)$ with $\Gamma_E = \{0.05\} \times (0, 1)$. Here a Dirichlet condition is imposed on $\{0\} \times (0, 1)$ so that the exact solution is defined by

$$u^{ex}(x, y) = \sum_{n=0}^{2L-1} \frac{1}{L} e^{i\mu_n x} \cos(n\pi y)$$

with $L = k/\pi$. Since the exact solution in this example includes a cutoff mode, it will also validate the effective performance of CRBC for cutoff modes.

It turns out that CRBC with $n_p = 2$ and $n_p = 3$ can reduce the reflection errors of 10 propagating modes up to 2.19494×10^{-4} and 2.29942×10^{-6}, respectively. As indicated by the reflection errors, Figure 6 shows that finite element approximations of CRBC with $n_p = 2$ converge at the quasi-optimal rate until the errors reach the level of the reflection error and those with $n_p = 3$ do not suffer from interruption of reflection error. In addition, since CRBC with the Neumann terminal condition provides the exact radiation condition for cutoff modes, we do not have any trouble to approximate solutions including a cutoff mode as seen in this experiment. In comparison to this result, it is shown in [17] that CRBC terminated with a Dirichlet condition gives the reflection error of cutoff modes proportional to

$$\left(\delta + \sum_{j=0}^{P} \frac{1}{a_j + \tilde{a}_j} \right)^{-1}$$

and hence it is required to employ small parameters to achieve a desirable accuracy, which in turn needs small finite element meshes.

In the next example, we consider the problem for $k = 10\pi + 10^{-6}$, where a near-cutoff mode of axial frequency $\mu_N \approx 0.0079$ with $N = 10$ is involved. In this case we choose a_0 and $\tilde{a}_0$ to be $-i\mu_N$ for removing the reflection of the

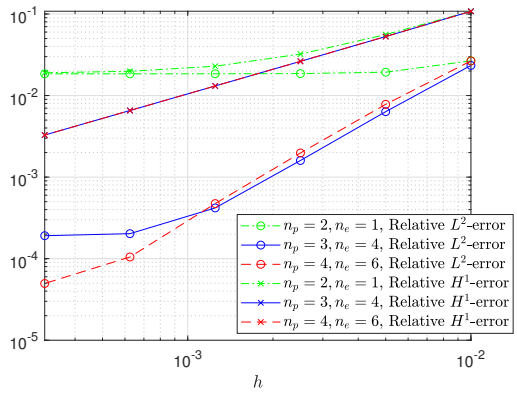


Figure 7: Relative errors of finite element approximations satisfying the CRBC for $h = 1/100, 1/200, 1/400, 1/800, 1/1600$ and $1/3200$ when $k = 10\pi + 10^{-6}$.

n_p	ρ_p	n_e	$e^{-\tilde{\mu}_{N+1}\delta}\rho_e$
2	2.09e-02	2	1.53e-03
3	2.19e-04	4	3.38e-05
4	2.30e-06	6	9.48e-07
5	2.41e-08	9	3.67e-09
6	2.52e-10	11	1.54e-10

(a) $\mu_{min} = \mu_{N-1} \approx 13.6939$

n_p	ρ_p	n_e	$e^{-\tilde{\mu}_{N+1}\delta}\rho_e$
2	2.60e-01	1	5.90e-03
3	9.36e-02	1	2.04e-02
4	3.37e-02	1	3.33e-02
5	1.22e-02	2	2.01e-03
6	4.38e-03	2	2.94e-03
7	1.58e-03	3	2.47e-04
8	5.69e-04	3	3.43e-04
9	2.05e-04	4	3.45e-05
10	7.40e-05	4	4.63e-05

(b) $\mu_{min} = \mu_N \approx 0.0079$

Table 4: Reflection errors of CRBC when $k = 10\pi + 10^{-6}$, $\delta = 0.05$ and $\tilde{\mu}_{min} = \tilde{\mu}_{N+1} \approx 14.3966$.

near-cutoff mode and use μ_{N-1} for the lower bound of the min-max problem (2.22). The reason why we set two values a_0 and $\tilde{a}_0$ to be equal is that otherwise the ratio $\tilde{a}_0/a_0$ would be so big, for example $\tilde{a}_0/a_0 > \mu_{N-1}/\mu_N \approx 1.72 \times 10^3$ as $\mu_{N-1} \approx 13.69$, that the bound of the second estimate of (5.15) would become large and render the stability constant of the auxiliary variables in (5.5) worse, which consequently would result in slow convergence. The resulting reflection errors of CRBC with $-i\mu_N$ as a parameter for each n_p are reported in Table 4a. When they are compared with the results given in Table 4b with μ_N used for the lower bound of the min-max problem 2.22, the former is apparently much smaller than the latter for each n_p. Figure 7 demonstrates the quasi-optimal convergence of finite element approximations until finite element errors are dominant.

7. Acknowledgment

This research of the author was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF-2018R1D1A1B07047416) funded by the Ministry of Education, Science and Technology.

References

[1] M. Abramowitz, I. A. Stegun, Handbook of mathematical functions with formulas, graphs, and mathematical tables, vol. 55, Dover, New York, 1964.

[2] N. I. Akhiezer, Elements of the theory of elliptic functions, American Mathematical Society, Providence, RI, 1990.

[3] W. Bangerth, R. Hartmann, G. Kanschat, deal.II—a general-purpose object-oriented finite element library, ACM Trans. Math. Software 33 (4) (2007) 24.

[4] E. Bécache, A.-S. Bonnet-Ben Dhia, G. Legendre, Perfectly matched layers for the convected Helmholtz equation, SIAM J. Numer. Anal. 42 (1) (2004) 409–433.

[5] B. Beckermann, A. Townsend, On the singular values of matrices with displacement structure, SIAM J. Matrix Anal. Appl. 38 (4) (2017) 1227–1248.

[6] A. Bendali, P. Guillaume, Non-reflecting boundary conditions for waveguides, Math. Comp. 68 (225) (1999) 123–144.

[7] D. Borisov, P. Exner, R. Gadyl′shin, D. Krejčiřík, Bound states in weakly deformed strips and layers, Ann. Henri Poincaré 2 (3) (2001) 553–572.

[8] R. Courant, D. Hilbert, Methods of Mathematical Physics, vol. 1, Wiley-Interscience, New York, 1953.

[9] E. B. Davies, L. Parnovski, Trapped modes in acoustic waveguides, Quart. J. Mech. Appl. Math. 51 (3) (1998) 477–492.

[10] V. Druskin, S. Güttel, L. Knizhnerman, Near-optimal perfectly matched layers for indefinite Helmholtz problems, SIAM Rev. 58 (1) (2016) 90–116.

[11] V. Druskin, L. Knizhnerman, Gaussian spectral rules for the three-point second differences. I. A two-point positive definite problem in a semi-infinite domain, SIAM J. Numer. Anal. 37 (2) (2000) 403–422 (electronic).

[12] V. Druskin, S. Moskow, Three-point finite-difference schemes, Padé and the spectral Galerkin method. I. One-sided impedance approximation, Math. Comp. 71 (239) (2002) 995–1019 (electronic).

[13] D. V. Evans, M. Levitin, D. Vassiliev, Existence theorems for trapped modes, J. Fluid Mech. 261 (1994) 21–31.

[14] P. Exner, P. Šeba, M. Tater, D. Vaněk, Bound states and scattering in quantum waveguides coupled laterally through a boundary window, J. Math. Phys. 37 (10) (1996) 4867–4887.

[15] P. Exner, S. A. Vugalter, Bound states in a locally deformed waveguide: the critical case, Lett. Math. Phys. 39 (1) (1997) 59–68.

[16] C. I. Goldstein, A finite element method for solving Helmholtz type equations in waveguides and other unbounded domains, Math. Comp. 39 (160) (1982) 309–324.

[17] T. Hagstrom, S. Kim, Complete radiation boundary conditions for the Helmholtz equation I: waveguides, Numer. Math. 141 (4) (2019) 917–966.

[18] I. Harari, I. Patlashenko, D. Givoli, Dirichlet-to-Neumann maps for unbounded wave guides, J. Comput. Phys. 143 (1) (1998) 200–223.

[19] R. L. Higdon, Absorbing boundary conditions for difference approximations to the multidimensional wave equation, Math. Comp. 47 (176) (1986) 437–459.

[20] R. L. Higdon, Numerical absorbing boundary conditions for the wave equation, Math. Comp. 49 (179) (1987) 65–90.

[21] D. Ingerman, V. Druskin, L. Knizherman, Optimal finite difference grids and rational approximations of the square root. I. Elliptic functions, Comm. Pure and Appl. Math. 53 (2000) 1039–1066.

[22] D. S. Jones, The eigenvalues of $\nabla^2 u + \lambda u = 0$ when the boundary conditions are given on semi-infinite domains, Proc. Cambridge Philos. Soc. 49 (1953) 668–684.

[23] S. Kim, Analysis of the convected Helmholtz equation with a uniform mean flow in a waveguide with complete radiation boundary conditions, J. Math. Anal. Appl. 410 (1) (2014) 275–291.

[24] S. Kim, Error analysis of PML-FEM approximations for the Helmholtz equation in waveguides, ESAIM Math. Model. Numer. Anal. 53 (4) (2019) 1191–1222.

[25] D. Krejčiřík, J. Kříz, On the spectrum of curved planar waveguides, Publ. Res. Inst. Math. Sci. 41 (3) (2005) 757–791.

[26] C. M. Linton, P. McIver, Embedded trapped modes in water waves and acoustics, Wave Motion 45 (1-2) (2007) 16–29.

[27] A. A. Medovikov, V. I. Lebedev, Variable time steps optimization of L_ω-stable Crank-Nicolson method, Russian J. Numer. Anal. Math. Modelling 20 (3) (2005) 283–303.

[28] D. A. Mitsoudis, C. Makridakis, M. Plexousakis, Helmholtz equation with artificial boundary conditions in a two-dimensional waveguide, SIAM J. Math. Anal. 44 (6) (2012) 4320–4344.

[29] P. Petrushev, V. Popov, Rational Approximation of Real Functions, vol. 28 of Encyclopedia of Mathematics, Cambridge University Press, Cambridge, 1987.