Noname manuscript No.
(will be inserted by the editor)

Convergence analysis of the continuous and discrete non-overlapping double sweep domain decomposition method based on PMLs for the Helmholtz equation

Seungil Kim · Hui Zhang

the date of receipt and acceptance should be inserted later

Abstract In this paper we will analyze the convergence of the non-overlapping double sweep domain decomposition method (DDM) with transmission conditions based on PMLs for the Helmholtz equation. The main goal is to establish the convergence of the double sweep DDM of both the continuous level problem and the corresponding finite element problem. We show that the double sweep process can be viewed as a contraction mapping of boundary data used for local subdomain problems not only in the continuous level and but also in the discrete level. It turns out that the contraction factor of the contraction mapping of the continuous level problem is given by an exponentially small factor determined by PML strength and PML width, whereas the counterpart of the discrete level problem is governed by the dominant term between the contraction factor similar to that of the continuous level problem and the maximal discrete reflection coefficient resulting from fast decaying evanescent modes. Based on this analysis we prove the convergence of approximate solutions in the H^1-norm. We also analyze how the discrete double sweep DDM depends on the number of subdomains and the PML parameters as the finite element discretization resolves sufficiently the Helmholtz and PML equations. Our theoretical results suggest that the contraction factor for the propagating modes depends linearly on the number of subdomains. To ensure the convergence, it is sufficient to have the PML width growing logarithmically with the number of subdomains. In the end, numerical experiments illustrating the convergence will be presented as well.

This research of the first 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.

S. Kim
Department of Mathematics and Research Institute for Basic Sciences, Kyung Hee University, Seoul, 02447, Korea
E-mail: sikim@khu.ac.kr

H. Zhang
Department of Applied Mathematics and Laboratory for Intelligent Computing and Financial Technology, Xi'an Jiaotong-Liverpool University (KSF-P-02, RDF-19-01-09)
Suzhou 215123, China
E-mail: hui.zhang@xjtlu.edu.cn

Keywords Double sweep domain decomposition · Helmholtz equation · Perfectly matched layer · Convergence analysis

Mathematics Subject Classification (2010) 65N12 · 65N30 · 65N55

1 Introduction

Solving time-harmonic wave propagation problems governed by the Helmholtz equation in the high-frequency regime leads to significant computational challenges due to the highly oscillatory nature of solutions that requires extremely large linear systems. For efficient solvers to the wave propagation problems various efforts have been made such as multigrid methods [6,13,37,40], shifted-Laplace preconditioners [9,14,16,39] and analytic incomplete LU (AILU) preconditioners [12,17,18].

The double sweep domain decomposition method (DDM) to be studied in this paper is conceptually originated from the idea of the last category among others. The AILU preconditioning technique in [17] is thought of as an approximate block LU factorization with a transparent Dirichlet-to-Neumann (DtN) condition for each subdomain problem replaced by a second order local approximation. Finding solutions by forward substitutions followed by backward substitutions sequentially with the approximate block LU factorization gives rise to an approximate inverse of the Helmholtz operator. Not surprisingly, its performance depends significantly on the accuracy of the approximate DtN condition. After the efficient absorbing boundary condition, so-called perfectly matched layer (PML), prevailed in wave propagation communities, the authors of [12] developed an improved preconditioner based on PML instead of the second order approximation in approximate LU factorization of [17].

A transmission condition based on a low-order approximate radiation condition was studied in [10] for Schwarz methods for the Helmholtz equation almost three decades ago. A quantitative convergence theory of the method of [10] for general domain and general decomposition was established only recently in [20]. The use of PML transmission conditions was first proposed in [38] and it was implemented successfully in [34]. The idea of utilizing PML for transmission conditions received more attention after the advent of [12] which brought about flourishing research on the double sweep DDM such as [8,15,29,35,41]. Each of these is developed from different points of views and has different formulations. The method of [8] transfers volume sources from one subdomain to a neighboring subdomain with radiation conditions on both sides of sweeping directions in the forward sweep whereas during the backward sweep it employs a radiation condition on the side to the direction of the backward sweep and a Dirichlet condition on the opposite side. On the other hand, subdomain problems of [29] communicate with neighboring subdomain problems via transmission conditions with Neumann data on interfaces instead of volume sources to reduce computational costs of subdomain problems. [35,41] also make use of Neumann data (which it is pointed out in [35] that have the same jump properties as single layer potentials) for subdomain problems but produce discontinuous approximate solutions compared with continuous approximate solutions of [8,29]. For an extensive review on sweeping preconditioners for the Helmholtz equation, we refer to [19] which gives unifying explanations of the equivalence between different formulations of sweeping domain decomposi-

tion solvers including AILU factorizations, source-transfer methods and optimized Schwarz methods.

It is worth noting that there are recent developments of the sweeping domain decomposition in two directions with special care of cross-points, rather than the one-way domain decomposition used in the above-mentioned methods, for solving the Helmholtz equation in the free space [30,32]. The former [30] is an extension of the source transfer method [8] based on PML transmission conditions, and the latter [32] proposes a new approach of an optimized Schwarz method utilizing Padé-type high-order absorbing transmission conditions.

In this paper, we consider a wave propagation problem governed by the Helmholtz equation posed in a waveguide. We take the double sweep formulation in [19] and employ PML with a piecewise constant coordinate stretching function studied in [25] for transmission conditions. As a main goal, we analyze convergence in both continuous and discrete levels. It is noted that convergence analyses of the double sweep DDM for continuous level problems can be found in [8,29]. In fact the double sweep DDM was applied in [33] almost two decades earlier to solve the convection-diffusion problem and its convergence was analyzed based on Fourier analysis, which is the main tool used in the convergence analysis of this paper. The methods in [8,29] generate a sequence of approximate solutions continuous on interfaces between subdomains, and a convergence of approximate solutions is proved in the H^1 Sobolev space for a full computational domain. However, the formulation in [19] gives rise to approximate solutions that are discontinuous on interfaces between subdomains. So we will provide an H^1-norm convergence of approximate solutions in each subdomain instead of the full domain, exhibiting that the continuous double sweep DDM converges linearly with contraction factor depending on the reflection coefficient of PML and the number of subdomains of a decomposition of the domain. Here we note that as opposed to the double sweep DDMs in [8,29], the method in this paper has the same formulation of each subdomain problem in the forward sweep and the backward sweep. Different formulations between the forward sweep and the backward sweep used in [8,29] are the price to pay for obtaining the continuity of each iterate.

On the other hand, to the extent of our knowledge, a convergence analysis for discrete level problems has not been yet available. To investigate the convergence of the double sweep DDM applied to the linear system resulting from the finite element discretization, we consider a quasi-uniform and shape-regular quadrilateral/hexahedral mesh of a computational domain, which is obtained by extruding a quasi-uniform and shape-regular mesh on a cross-sectional boundary into the waveguide with uniformly distributed grids along the axis of the waveguide. With this specialized mesh, we can find local solutions to subdomain problems by using the dispersion analysis for each discrete cross-sectional eigenvalue. With the help of the solution formula in terms of the roots of the characteristic equation of certain difference equations, discrete DtN operators for the exact radiation condition and PML can be derived. Surprisingly, it turns out that the reflection coefficients of the discrete level problem behave differently from those of the continuous level problem. More precisely, whereas the reflection coefficients of the continuous level problem decay exponentially as the decay rate of evanescent modes increases, those of the discrete problem grow to some number close to one in magnitude for large mode indices. As a result, if fast decaying evanescent modes are involved in iterates in the discrete level, sequences generated by the double sweep DDM are doomed

to converge slowly, although we can expect the same rate of convergence of the discrete problems as that of the continuous problem otherwise.

As for the convergence rate of the method, there are many significant factors determining the maximal active discrete reflection coefficient in numerical implementations such as what types of Fourier modes compose each sequence generated by the double sweep algorithm, as mentioned earlier, and how strong PML parameters are used. The performance of the discrete double sweep DDM depends on wavenumber as well, however it is in the sense of that it depends on the location of wavenumber with respect to the distribution of the cross-sectional eigenvalues. Indeed, the closer the wavenumber is to cross-sectional eigenvalues the larger the discrete maximal reflection coefficient is. More importantly, the main convergence analysis reveals that the convergence rate is also affected by the number of subdomains, denoted by J, that is, as the finite element discretization resolves sufficiently the Helmholtz and PML equations, the contraction factor for the propagating modes depends linearly on the number of subdomains, J, and hence the number of iterations can increase proportionally to the log of J. From a computational point of view, it is important to minimize computational costs by taking small number of grid points in PML along the axis of waveguides, denoted by N_p. It is shown that the growth of N_p is of $O(\ln(J))$ as well as $O(\ln(n))$ to keep the same number of iterations with n grid points in each direction of square domains with uniform mesh, and the cost required for one sweep is of order $O(N_p^2 n^2)$. This result is consistent with one observed in [12,35] based on PML of piecewise quadratic stretching functions. These parameter-dependence of the double sweep DDM will be advocated by various numerical experiments.

At last, this paper is organized as follows. In Section 2 for preliminaries, we discuss the solution formula of the Helmholtz equation in a straight waveguide and study PML as an approximate DtN operator. In Section 3, we introduce the double sweep DDM algorithm in the continuous level and analyze its convergence. In Section 4, we study the finite element subdomain transmission problem. Here we discuss how solutions of the finite element problem in the reference subdomain are represented in terms of discrete cross-sectional eigenvectors and how errors propagate in local subdomains by analyzing discrete transmission conditions. Section 5 deals with the double sweep DDM algorithm in the discrete level and presents the convergence analysis of the discretized problem. Section 6 is devoted to showing numerical experiments to illustrate the convergence theory.

2 Preliminaries

2.1 Model problem

Let $\Omega_\infty = (-\infty, a) \times \Gamma$ be a semi-infinite waveguide in $\mathbb{R}^{\mathrm{d}}$, $\mathrm{d} = 2$ or 3 with $a > 0$ constant and Γ a Lipschitz bounded domain in $\mathbb{R}^{\mathrm{d}-1}$. See Fig. 1. We assume that the axis of the waveguide Ω_∞ is parallel to the x-axis for $(x, y) \in \mathbb{R} \times \mathbb{R}^{\mathrm{d}-1}$. As a model problem, we consider the Helmholtz equation with positive wavenumber k and a wave source f in $L^2(\Omega_\infty)$ supported for $x > 0$,

$$\begin{aligned} -\Delta u - k^2 u &= f \text{ in } \Omega, \\ \frac{\partial u}{\partial \nu} &= 0 \text{ on } \partial\Omega \setminus \bar{\Gamma}_0 \quad \text{and} \quad \frac{\partial u}{\partial \nu} = T(u) \text{ on } \Gamma_0, \end{aligned} \tag{2.1}$$

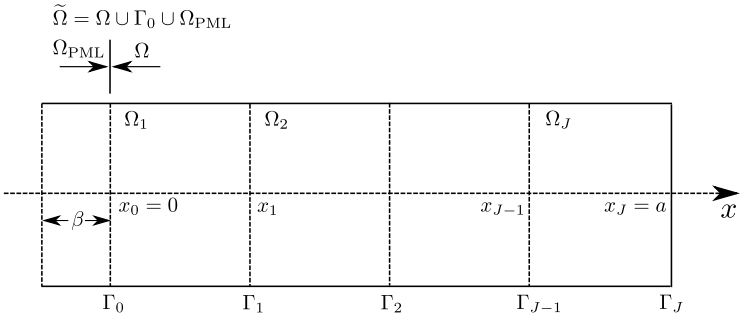


Fig. 1: Semi-infinite waveguide and non-overlapped decomposition of the domain

where Ω is a Lipschitz bounded domain obtained by truncating Ω_∞ at $x = 0$ with $\Gamma_0 := \{0\} \times \Gamma$ and $\Gamma_J := \{J\} \times \Gamma$ the cross-sectional boundaries at $x = 0$ and $x = a$, respectively, which can be identified with Γ. Also ν stands for the unit normal vector pointing outward from the domain and T is the DtN operator for the radiation condition at infinity.

We assume that k^2 is not an eigenvalue of the problem (2.1) and in addition that there is no cutoff mode for the solvability of local subdomain problems, that is, if λ_n^2 denotes Neumann eigenvalues ordered increasingly, $0 = \lambda_0^2 \le \lambda_1^2 \le \cdots$, of the negative cross-sectional Laplace operator

$$\begin{aligned} -\Delta_y Y_n &= \lambda_n^2 Y_n \text{ in } \Gamma, \\ \frac{\partial Y_n}{\partial \nu} &= 0 \text{ on } \partial\Gamma, \end{aligned} \tag{2.2}$$

then $\lambda_n \neq k$ for all n. We take an orthonormal basis $\{Y_n\}_{n=0}^\infty$ in $L^2(\Gamma)$ consisting of Neumann eigenfunctions and so general solutions to the Helmholtz equation in the straight waveguide for $0 < x < a$ can be written as

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

where $\mu_n = \sqrt{k^2 - \lambda_n^2}$ with the branch cut of the negative real axis. Since λ_n tends toward infinity as $n \to \infty$, there exists $N > 0$ (we call it the cutoff index) such that $\lambda_n < k$ for $n \le N$ and $\lambda_n > k$ for $n > N$ so that $\mu_n > 0$ for $n \le N$ represents the axial frequency or wavenumber of propagating modes and the imaginary part $\tilde{\mu}_n > 0$ of $\mu_n = i\tilde{\mu}_n$ for $n > N$ is the decay rate of evanescent modes. For norm estimates associated with traces on cross-sectional boundaries identified with Γ, we introduce Sobolev spaces $H^s(\Gamma)$ for $-1 \le s \le 1$ of functions $\phi = \sum_{n=0}^\infty \phi_n Y_n$ such that

$$\|\phi\|_{H^s(\Gamma)}^2 := \sum_{n=0}^{\infty} (1 + \lambda_n^2)^s |\phi_n|^2 < \infty.$$

see e.g., [27]. Here $H^{-s}(\Gamma)$ is understood as the dual space of $H^s(\Gamma)$ for $0 \le s \le 1$ and we use $\langle \cdot, \cdot \rangle_{s,\Gamma}$ for the duality pairing between $H^s(\Gamma)$ and $H^{-s}(\Gamma)$ for $0 \le s \le 1$.

For numerical computations, the exact radiation condition in (2.1) based on the DtN operator $T : H^{1/2}(\Gamma_0) \to H^{-1/2}(\Gamma_0)$,

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

for $\phi = \sum_{n=0}^{\infty} \phi_n Y_n$ in $H^{1/2}(\Gamma_0)$, needs to be replaced by an accurate absorbing boundary condition such as PML [25], complete radiation boundary condition (CRBC) [21,26] or rational approximation [11] for the DtN operator. In this paper we will employ PML not only for absorbing boundary conditions of the global problem (2.1) but also for transmission conditions of local subdomain problems.

2.2 PML operators T_{PML}

In this section we review basic theories about PML in waveguides (see e.g., [25]). The PML used for the model problem is defined in terms of a piecewise constant coordinate stretching function

$$\tilde{x} = \begin{cases} (\sigma_r + i\sigma_i)x & \text{for } x < 0, \\ x & \text{for } x > 0, \end{cases} \qquad \sigma = \frac{\mathrm{d}\tilde{x}}{\mathrm{d}x} = \begin{cases} \sigma_0 = \sigma_r + i\sigma_i & \text{for } x < 0, \\ 1 & \text{for } x > 0 \end{cases}$$

with positive constants σ_r, σ_i and is terminated with a homogeneous Neumann condition on $\Gamma_{\mathrm{PML}} = \{-\beta\} \times \Gamma$, where β is a parameter for the PML width. The parameters σ_r and σ_i are determined in such a way that the smallest decay rate of converted modes from propagating modes and the smallest decay rate of those from evanescent modes are balanced,

$$\mu_N \sigma_i = \tilde{\mu}_{N+1} \sigma_r := \sigma_\mu, \tag{2.3}$$

which is called the PML strength. These relations determine $\sigma_0 = \frac{\sigma_\mu}{\tilde{\mu}_{N+1}} + i\frac{\sigma_\mu}{\mu_N}$. The reason that the Neumann condition is used for the terminal condition on Γ_{PML} instead of the usual homogeneous Dirichlet condition is that PML with the Neumann condition has the better performance in waveguides than PML with the Dirichlet condition in case that near-cutoff modes (modes corresponding to $\mu_{\min} := \min\{\mu_N, \tilde{\mu}_{N+1}\} \ll 1$) exist, see [26].

Denoting $\widetilde{\Omega} = \Omega \cup \Gamma_0 \cup \Omega_{\mathrm{PML}}$ with $\Omega_{\mathrm{PML}} = (-\beta, 0) \times \Gamma$ (see Fig. 1), the variational problem for the PML Helmholtz equation can be written as finding $\tilde{u} \in H^1(\widetilde{\Omega})$ satisfying

$$\tilde{b}(\tilde{u}, \tilde{v}) = (f, \tilde{v})_\Omega \text{ for } \tilde{v} \in H^1(\widetilde{\Omega}), \tag{2.4}$$

where

$$\tilde{b}(\tilde{u}, \tilde{v}) = (H\nabla\tilde{u}, \nabla\tilde{v})_{\widetilde{\Omega}} - k^2(\sigma\tilde{u}, \tilde{v})_{\widetilde{\Omega}}$$

with $H = diag(\sigma^{-1}, \sigma I_{\mathrm{d}-1})$. Here $(\cdot,\cdot)_{\mathcal{D}}$ is the L^2-inner product in the domain $\mathcal{D}$. On the other hand, the sesquilinear form associated with PML on Ω_{PML}

$$b_{\mathrm{PML}}(u, v) = (\frac{1}{\sigma}\frac{\partial u}{\partial x}, \frac{\partial v}{\partial x})_{\Omega_{\mathrm{PML}}} + (\sigma_0 \nabla_y u, \nabla_y u)_{\Omega_{\mathrm{PML}}} - k^2(\sigma_0 u, v)_{\Omega_{\mathrm{PML}}}$$

satisfies a coercivity (see [25]) in $\widetilde{H}_0^1(\Omega_{\rm PML})$, the space of functions in $H^1(\Omega_{\rm PML})$ vanishing on Γ_0, which allows us to define a continuous extension operator $S : H^{1/2}(\Gamma_0) \to H^1(\Omega_{\rm PML})$ by $S(g) = u$, where u is the unique solution to the problem

$$b_{\rm PML}(u, v) = 0 \text{ for } v \in \widetilde{H}_0^1(\Omega_{\rm PML}) \tag{2.5}$$

with $u = g$ on Γ_0. Then the PML operator $T_{\rm PML} : H^{1/2}(\Gamma_0) \to H^{-1/2}(\Gamma_0)$ is defined as a DtN operator by $T_{\rm PML}(g) = -\frac{1}{\sigma_0}\frac{\partial S(g)}{\partial x}$ for $g \in H^{1/2}(\Gamma_0)$. The PML operator $T_{\rm PML}$ can also be thought of as a variational normal derivative,

$$\langle T_{\rm PML}(g), v\rangle_{\Gamma_0} = -b_{\rm PML}(S(g), S(v)) \text{ for } v \in H^{1/2}(\Gamma_0). \tag{2.6}$$

Now, we consider a problem to find $u \in H^1(\Omega)$ satisfying

$$b(u, v) = (f, v)_\Omega \text{ for } v \in H^1(\Omega), \tag{2.7}$$

where

$$b(u, v) = (\nabla u, \nabla v)_\Omega - k^2(u, v)_\Omega - \langle T_{\rm PML}(u), v\rangle_{\Gamma_0}. \tag{2.8}$$

It is also shown in [25] that this problem (2.7) is well-posed and equivalent to the problem (2.4), that is, the restriction of the solution $\tilde{u}$ of the problem (2.4) to Ω coincides with the solution u to the problem (2.7). Its proof proceeds by using the fact that $T_{\rm PML}$ is an approximation of T. Indeed, the Fourier analysis yields that the PML operator $T_{\rm PML}$ can be expressed as

$$T_{\rm PML}(\phi) = \sum_{n=1}^{\infty} \left(i\mu_n \frac{1 - e^{2i\sigma_0\mu_n\beta}}{1 + e^{2i\sigma_0\mu_n\beta}} \right) \phi_n Y_n := \sum_{n=1}^{\infty} \Lambda_n \phi_n Y_n \tag{2.9}$$

and hence the convergence of $T_{\rm PML}$ to T can be proved in the sense that there exists a positive constant M such that for $\sigma_\mu\beta > M$

$$\|(T - T_{\rm PML})(\phi)\|_{H^{-1/2}(\Gamma_0)} \lesssim e^{-2\sigma_\mu\beta}\|\phi\|_{H^{1/2}(\Gamma_0)}.$$

From here on, we will use $a \lesssim b$ for $a \le Cb$ with a generic constant C that may depend on Ω and k but is independent of PML parameters and the number of subdomains of the double sweep DDM as well as functions involved in estimates.

Remark 2.1 *The* PML *of width β and complex coordinate stretching constant σ_0 can be applied to any cross-sectional boundary identical with Γ not only on Γ_0. In addition, we will use it for transmission conditions on cross-sectional interfaces between subdomains throughout the paper. Also we can use* PML *with different parameters β and σ_0 for different interfaces, however we will take one* PML *for a simple presentation.*

Remark 2.2 *Although* PML *with a piecewise constant coordinate stretching function is used in this paper for absorbing and transmission conditions of the double sweep DDM, there are many other types of* PML *coordinate stretching functions such as piecewise polynomials [5, 7, 12, 28, 36] and unbounded functions [3, 42]. In particular, it is shown in [36] that* PML *equipped with a piecewise quadratic polynomial stretching function normalized with respect to k and* PML *width is utilized for an efficient transmission condition of the double sweep DDM. It turns out that* PML *of a piecewise quadratic polynomial stretching function gives a better performance than the piecewise constant counterpart, however the convergence analysis for the double sweep DDM with* PML *of the quadratic stretching function is much more difficult in the discrete level and we leave it for the future research.*

3 Double sweep for the continuous problem

The domain Ω is decomposed non-overlappingly in one-way along the axis of the waveguide, the x-axis for $(x, y) \in \mathbb{R} \times \mathbb{R}^{\mathrm{d}-1}$, $\Omega = \cup_{j=1}^{J} \Omega_j$ with

$$\Omega_j = (x_{j-1}, x_j) \times \Gamma \quad \text{for } j = 1, \ldots, J.$$

Here $0 = x_0 < x_1 < \ldots < x_{J-1} < a = x_J$. See Fig. 1. For a simple presentation, we assume that x_j are evenly spaced so that $H := x_j - x_{j-1}$ is constant. We denote interfaces between two neighboring subdomains by $\Gamma_j = \{x_j\} \times \Gamma$ for $j = 1, 2, \ldots, J-1$.

We note that if u^{ex} is the solution to the problem (2.7), then the restriction $u_j^{ex} = u^{ex}|_{\Omega_j}$ is the unique solution to the local subdomain problem

$$-\Delta u_j^{ex} - k^2 u_j^{ex} = f_j \text{ in } \Omega_j, \tag{3.1}$$

$$\left(\frac{\partial}{\partial \nu_j} - T_{\mathrm{PML}}\right) u_j^{ex} = \left(\frac{\partial}{\partial \nu_j} - T_{\mathrm{PML}}\right) u_{j-1}^{ex} \text{ on } \Gamma_{j-1}, \tag{3.2}$$

$$\left(\frac{\partial}{\partial \nu_j} - T_{\mathrm{PML}}\right) u_j^{ex} = \left(\frac{\partial}{\partial \nu_j} - T_{\mathrm{PML}}\right) u_{j+1}^{ex} \text{ on } \Gamma_j \tag{3.3}$$

with a homogeneous Neumann condition imposed on boundaries other than Γ_j with $j = 1, \ldots, J$, where ν_j stands for the outward unit normal vector on $\partial\Omega_j$ and f_j is the restriction of f to Ω_j. The right-hand-side of (3.2) for $j = 1$ is set to be zero, and the transmission condition (3.3) for $j = J$ is replaced with a homogeneous Neumann condition on Γ_J. The local subdomain problem can also be written as a variational form,

$$b_j(u_j^{ex}, v) = (f_j, v)_{\Omega_j} + \langle \gamma_j^{\mathrm{L}}, v \rangle_{\Gamma_{j-1}} + \langle \gamma_j^{\mathrm{R}}, v \rangle_{\Gamma_j} \text{ for all } v \in V_{\Omega_j} := H^1(\Omega_j), \tag{3.4}$$

where $b_j(\cdot, \cdot)$ is the sesquilinear form in $V_{\Omega_j} \times V_{\Omega_j}$ analogous to (2.8) defined by

$$b_j(u, v) = (\nabla u, \nabla v)_{\Omega_j} - k^2 (u, v)_{\Omega_j} - \langle T_{\mathrm{PML}}(u), v \rangle_{\Gamma_{j-1} \cup \Gamma_j} \quad \text{for } u, v \in V_{\Omega_j}$$

and

$$\gamma_j^{\mathrm{L}} = \left(\frac{\partial}{\partial \nu_j} - T_{\mathrm{PML}}\right) u_{j-1}^{ex}, \qquad \gamma_j^{\mathrm{R}} = \left(\frac{\partial}{\partial \nu_j} - T_{\mathrm{PML}}\right) u_{j+1}^{ex} \tag{3.5}$$

for $j = 1, \ldots, J$ with the obvious modifications for $j = 1$ and $j = J$. It is shown in [25] that for any $f_j \in L^2(\Omega_j)$ and any $\gamma_j^{\mathrm{L}} \in H^{-1/2}(\Gamma_{j-1})$, $\gamma_j^{\mathrm{R}} \in H^{-1/2}(\Gamma_j)$ the problem (3.4) has a unique solution $u_j \in H^1(\Omega_j)$ satisfying

$$\|u_j\|_{H^1(\Omega_j)} \lesssim \|f_j\|_{L^2(\Omega_j)} + \|\gamma_j^{\mathrm{L}}\|_{H^{-1/2}(\Gamma_{j-1})} + \|\gamma_j^{\mathrm{R}}\|_{H^{-1/2}(\Gamma_j)}.$$

Since the boundary data (3.5) are unknown in seeking for the solution u^{ex} in practice, an iterative method by using approximate data obtained from previous iterates in neighboring subdomains can be a reasonable alternative, which leads to the double sweeping iterative solver introduced in [19]. The formulations in this paper for the double sweep iteration are the same as those in [19] and we provide a convergence analysis for the double sweep DDM in both continuous and discrete levels.

For the convergence analysis in the continuous level, we let $\boldsymbol{V} = \prod_{j=1}^{J} V_{\Omega_j}$ equipped with the norm

$$\|\boldsymbol{u}\|_{\boldsymbol{V}} = \left(\sum_{j=1}^{J} \|u_j\|^2_{H^1(\Omega_j)} \right)^{1/2}$$

for $\boldsymbol{u} = (u_1, u_2, \ldots, u_J) \in \boldsymbol{V}$. The double sweep DDM yields an approximation $\boldsymbol{u} = (u_1, u_2, \ldots, u_J)$ in $\boldsymbol{V}$ to the solution $\boldsymbol{u}^{ex} = (u_1^{ex}, u_2^{ex}, \ldots, u_J^{ex})$.

3.1 Algorithm

To obtain an approximation to $\boldsymbol{u}^{ex}$, with any initial iterate $\boldsymbol{u}^0 = (u_1^0, \ldots, u_J^0) \in \boldsymbol{V}$ such that $\frac{\partial u_j^0}{\partial \nu_j} \in H^{-1/2}(\Gamma_{j-1} \cup \Gamma_j)$, we find the m-th iterate $\boldsymbol{u}^m$ for $m = 1, 2, \ldots$ by solving sequentially subdomain problems in Ω_j from $j = 1$ to $j = J-1$ in the forward sweep and then solving subdomain problems from $j = J$ to $j = 1$ in the backward sweep. In solving each local subdomain problem involved in the forward and backward sweeps, we take three steps as follows: for a current iterate $\boldsymbol{u} \in \boldsymbol{V}$

1. Extract data γ_j^{L} and γ_j^{R} coming into Ω_j from $\boldsymbol{u}$.
2. Solve the local problem for $\phi_j \in V_{\Omega_j}$ satisfying

$$b_j(\phi_j, v) = (f_j, v)_{\Omega_j} + \langle \gamma_j^{\mathrm{L}}, v \rangle_{\Gamma_{j-1}} + \langle \gamma_j^{\mathrm{R}}, v \rangle_{\Gamma_j} \text{ for all } v \in V_{\Omega_j}. \tag{3.6}$$

3. Update the j-th component of $\boldsymbol{u}$ with ϕ_j.

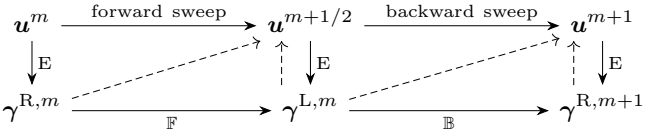
Remark 3.1 *When we extract the incoming data, we observe that*

(1) At the beginning of the forward sweep in the $(m+1)$-th iteration, all $\gamma_j^{\mathrm{R},m}$ for $j = 1, 2, \ldots, J-1$ that will be used for the forward sweep can be computed from $\boldsymbol{u}^m$. Thus we can denote the boundary data coming into each subdomain from the right used for the forward sweep in the $(m+1)$-th iteration by $\boldsymbol{\gamma}^{\mathrm{R},m} = (\gamma_1^{\mathrm{R},m}, \gamma_2^{\mathrm{R},m}, \ldots, \gamma_{J-1}^{\mathrm{R},m})$ in $G^{\mathrm{R}} := \prod_{j=1}^{J-1} H^{-1/2}(\Gamma_j)$ without solving intermediate problems of the forward sweep. However, $\gamma_j^{\mathrm{L},m}$ is updated immediately after the local problem in Ω_{j-1} is solved in the forward sweep.

(2) Denoting the intermediate iterate by $\boldsymbol{u}^{m+1/2}$ after finishing the forward sweep of the $(m+1)$-th iteration, all $\gamma_j^{\mathrm{L},m}$ for $j = 2, \ldots, J$ that will be used for the backward sweep can be computed from $\boldsymbol{u}^{m+1/2}$. Thus we can denote the boundary data coming into each subdomain from the left used for the backward sweep in the $(m+1)$-th iteration by $\boldsymbol{\gamma}^{\mathrm{L},m} = (\gamma_2^{\mathrm{L},m}, \gamma_3^{\mathrm{L},m}, \ldots, \gamma_J^{\mathrm{L},m})$ in $G^{\mathrm{L}} := \prod_{j=2}^{J} H^{-1/2}(\Gamma_j)$ without solving intermediate problems of the backward sweep. However, $\gamma_j^{\mathrm{R},m+1}$ is updated immediately after the local problem in Ω_{j+1} is solved in the backward sweep.

(3) From (1), we can view the forward sweep of the $(m+1)$-th iteration as a mapping $\mathbb{F} : \boldsymbol{\gamma}^{\mathrm{R},m} \mapsto \boldsymbol{\gamma}^{\mathrm{L},m}$. Also, from (2), the backward sweep of the $(m+1)$-th iteration can be regarded as a mapping $\mathbb{B} : \boldsymbol{\gamma}^{\mathrm{L},m} \mapsto \boldsymbol{\gamma}^{\mathrm{R},m+1}$. See the following diagram of the $(m+1)$-th iteration, where E *stands for extraction of incoming data. In addition,*

$\boldsymbol{u}^{m+1/2}$ can be computed from $\boldsymbol{\gamma}^{\mathrm{L},m}$ and $\boldsymbol{\gamma}^{\mathrm{R},m}$ along the dashed arrows in the left part of the diagram. Similarly, $\boldsymbol{u}^{m+1}$ can be computed from $\boldsymbol{\gamma}^{\mathrm{L},m}$ and $\boldsymbol{\gamma}^{\mathrm{R},m+1}$.



(4) The data $\gamma_j^{\mathrm{L},m}$ used for the problem in Ω_j in the forward sweep ($2 \le j \le J-1$) is used again for the problem in Ω_j in the backward sweep of the same iteration.

(5) The data $\gamma_j^{\mathrm{R},m+1}$ used for the problem in Ω_j in the backward sweep ($1 \le j \le J-1$) is used again for the problem in Ω_j in the forward sweep of the next iteration.

The algorithm of the double sweep DDM for the continuous level is presented in Algorithm 1 with the unnecessary fractional index $m+1/2$ removed, i.e., $\boldsymbol{u}^{m+1/2}$ after the forward sweep is considered as $\boldsymbol{u}^{m+1}$ in the middle of $(m+1)$-th iteration. Since the first subdomain problem in Ω_1 of the forward sweep in the next iteration is also considered as the last subdomain problem of the backward sweep in the current iteration, every subdomain problem in Ω_j needs to be solved twice except in Ω_1 and Ω_J during one double sweep. Thus, one iteration of the double sweep DDM requires $2(J-1)$ local subdomain problems in the actual practice.

Since it is essential to understand how the data $\gamma_j^{\mathrm{L},m}$ and $\gamma_j^{\mathrm{R},m}$ coming into Ω_j are transferred to neighboring subdomains $\Omega_{j\pm1}$ after solving the local problem in Ω_j, we will investigate a transmission problem in a reference domain identical with Ω_j in the next subsection.

Algorithm 1 Double Sweep DDM for the continuous level problem

Choose any initial iterate $\boldsymbol{u}^0 = (u_1^0, u_2^0, \ldots, u_J^0)$ satisfying the regularity

$$\left.\frac{\partial u_j^0}{\partial \nu_j}\right|_{\Gamma_{j-1}} \in H^{-1/2}(\Gamma_{j-1}) \quad \text{and} \quad \left.\frac{\partial u_j^0}{\partial \nu_j}\right|_{\Gamma_j} \in H^{-1/2}(\Gamma_j). \tag{3.7}$$

Compute $\boldsymbol{\gamma}^{\mathrm{R},0} = (\gamma_1^{\mathrm{R},0}, \ldots, \gamma_{J-1}^{\mathrm{R},0})$ from $\boldsymbol{u}^0$.
Set $m = 0$.
while the residual is larger than given tolerance **do**
  Set $\boldsymbol{u}^{m+1} \leftarrow \boldsymbol{u}^m$. ▷ Forward sweep
  for $j = 1, \ldots, J-1$ **do**
    i. Compute $\gamma_j^{\mathrm{L},m}$ from u_{j-1}^{m+1} with $\gamma_1^{\mathrm{L},m} = 0$.
    ii. Solve the local problem (3.6) with $\gamma_j^{\mathrm{L}} = \gamma_j^{\mathrm{L},m}$ and $\gamma_j^{\mathrm{R}} = \gamma_j^{\mathrm{R},m}$ for $\phi_j \in V_{\Omega_j}$.
    iii. Update the j-th component u_j^{m+1} of $\boldsymbol{u}^{m+1}$ by using ϕ_j.
  end for
  Compute $\gamma_J^{\mathrm{L},m}$ from u_{J-1}^{m+1}. ▷ Backward sweep
  for $j = J, J-1, \ldots, 1$ **do**
    i. Compute $\gamma_j^{\mathrm{R},m+1}$ from u_{j+1}^{m+1} with $\gamma_J^{\mathrm{R},m+1}$ being ignored.
    ii. Solve the local problem (3.6) with $\gamma_j^{\mathrm{L}} = \gamma_j^{\mathrm{L},m}$ and $\gamma_j^{\mathrm{R}} = \gamma_j^{\mathrm{R},m+1}$ for $\phi_j \in V_{\Omega_j}$.
    iii. Update the j-th component u_j^{m+1} of $\boldsymbol{u}^{m+1}$ by using ϕ_j.
  end for
  Set $m \leftarrow m+1$.
end while

3.2 Subdomain transmission problem

Let $\hat{\Omega}$ be a reference waveguide $\hat{\Omega} = (\mathrm{L}, \mathrm{R}) \times \Gamma$ of width $H = \mathrm{R} - \mathrm{L} > 0$ with boundaries $\Gamma_{\mathrm{L}} = \{\mathrm{L}\} \times \Gamma$ and $\Gamma_{\mathrm{R}} = \{\mathrm{R}\} \times \Gamma$. We consider the wave propagation problem to find $u \in H^1(\hat{\Omega})$ satisfying

$$\Delta u + k^2 u = 0 \text{ in } \hat{\Omega}, \tag{3.8}$$

$$\frac{\partial u}{\partial \nu} = T_{\mathrm{PML}}(u) + \gamma_{\mathrm{in}}^{\mathrm{L}} \text{ on } \Gamma_{\mathrm{L}}, \qquad \frac{\partial u}{\partial \nu} = T_{\mathrm{PML}}(u) + \gamma_{\mathrm{in}}^{\mathrm{R}} \text{ on } \Gamma_{\mathrm{R}} \tag{3.9}$$

together with $\partial u/\partial \nu = 0$ on $\partial\hat{\Omega} \setminus \overline{\Gamma_{\mathrm{L}}} \cup \overline{\Gamma_{\mathrm{R}}}$, where $\gamma_{\mathrm{in}}^{\mathrm{L}} = \sum_{n=1}^{\infty} \gamma_{\mathrm{in},n}^{\mathrm{L}} Y_n \in H^{-1/2}(\Gamma_{\mathrm{L}})$ and $\gamma_{\mathrm{in}}^{\mathrm{R}} = \sum_{n=1}^{\infty} \gamma_{\mathrm{in},n}^{\mathrm{R}} Y_n \in H^{-1/2}(\Gamma_{\mathrm{R}})$ are two input data coming into the domain $\hat{\Omega}$ through the boundaries $\Gamma_{\mathrm{L}} \cup \Gamma_{\mathrm{R}}$. Once solving the problem, we will find the outgoing data $\gamma_{\mathrm{out}}^{\mathrm{L}}$ and $\gamma_{\mathrm{out}}^{\mathrm{R}}$ of the solution u,

$$\gamma_{\mathrm{out}}^{\mathrm{L}} := -\frac{\partial u}{\partial \nu} - T_{\mathrm{PML}}(u) \text{ on } \Gamma_{\mathrm{L}} \quad \text{and} \quad \gamma_{\mathrm{out}}^{\mathrm{R}} := -\frac{\partial u}{\partial \nu} - T_{\mathrm{PML}}(u) \text{ on } \Gamma_{\mathrm{R}}. \tag{3.10}$$

We recall that the PML operator $T_{\mathrm{PML}} : H^{1/2}(\Gamma) \to H^{-1/2}(\Gamma)$ is defined by (2.9) with the identification between Γ and $\Gamma_{\mathrm{L/R}}$. The reflection coefficient of the n-th mode associated with the PML operator is given by

$$Q_n = \frac{i\mu_n - \Lambda_n}{-i\mu_n - \Lambda_n} = e^{2i\mu_n \sigma_0 \beta},$$

which is bounded by

$$|Q_n| \le e^{-2\sigma_\mu \beta} \text{ for all } n = 0, 1, \ldots \tag{3.11}$$

due to (2.3). Here we assume that the PML parameters σ_μ and β are chosen so that $e^{-2\sigma_\mu \beta} < 1/2$.

We begin by examining the coefficient of the n-th mode of the solution u to the problem (3.8)-(3.9), written as

$$u_n = A_n e^{i\mu_n x} + B_n e^{-i\mu_n x}.$$

Two boundary conditions (3.9) lead to

$$\begin{aligned}
(-i\mu_n - \Lambda_n) e^{i\mu_n \mathrm{L}} A_n + (i\mu_n - \Lambda_n) e^{-i\mu_n \mathrm{L}} B_n &= \gamma_{\mathrm{in},n}^{\mathrm{L}} \text{ on } \Gamma_{\mathrm{L}}, \\
(i\mu_n - \Lambda_n) e^{i\mu_n \mathrm{R}} A_n + (-i\mu_n - \Lambda_n) e^{-i\mu_n \mathrm{R}} B_n &= \gamma_{\mathrm{in},n}^{\mathrm{R}} \text{ on } \Gamma_{\mathrm{R}}.
\end{aligned}$$

The solution to the linear system is given by

$$\begin{bmatrix} A_n \\ B_n \end{bmatrix} = \frac{e^{i\mu_n H}}{(1 - Q_n^2 e^{2i\mu_n H})(-i\mu_n - \Lambda_n)} \begin{bmatrix} e^{-i\mu_n \mathrm{R}} & -Q_n e^{-i\mu_n \mathrm{L}} \\ -Q_n e^{i\mu_n \mathrm{R}} & e^{i\mu_n \mathrm{L}} \end{bmatrix} \begin{bmatrix} \gamma_{\mathrm{in},n}^{\mathrm{L}} \\ \gamma_{\mathrm{in},n}^{\mathrm{R}} \end{bmatrix}.$$

From (3.10) it can be obtained by a straightforward computation that

$$\begin{bmatrix} \gamma_{\mathrm{out},n}^{\mathrm{L}} \\ \gamma_{\mathrm{out},n}^{\mathrm{R}} \end{bmatrix} = \begin{bmatrix} \varepsilon_n & \zeta_n \\ \zeta_n & \varepsilon_n \end{bmatrix} \begin{bmatrix} \gamma_{\mathrm{in},n}^{\mathrm{L}} \\ \gamma_{\mathrm{in},n}^{\mathrm{R}} \end{bmatrix}, \tag{3.12}$$

where

$$\zeta_n = \frac{(1 - Q_n^2) e^{i\mu_n H}}{1 - Q_n^2 e^{2i\mu_n H}}, \quad \varepsilon_n = \frac{(1 - e_n^{2i\mu_n H}) Q_n}{1 - Q_n^2 e^{2i\mu_n H}}.$$

According to the formula (3.12), ε_n and ζ_n can be interpreted as the coefficients that measure how much of the incident fields coming through one side boundary are reflected back out to the same side boundary and are propagating out to the other side boundary, respectively. Also, it is seen that ε_n is asymptotically bounded by the reflection coefficient Q_n. On the other hand, ζ_n represents the phase change approximately of $e^{i\mu_n H}$ for propagation modes and shows the amount of decay for evanescent modes while the modes are traveling from one side to the other.

Lemma 3.2 *It holds that*

$$\begin{aligned} |\varepsilon_n| &< 3e^{-2\sigma_\mu \beta}, \\ |\zeta_n| &< 1 + 3e^{-4\sigma_\mu \beta}. \end{aligned} \tag{3.13}$$

Proof Using (3.11) and noting that $|e^{i\mu_n H}| \le 1$, we prove that

$$\begin{aligned} |\varepsilon_n| &= \left| \frac{(1 - e^{2i\mu_n H})Q_n}{1 - Q_n^2 e^{2i\mu_n H}} \right| \le \frac{2}{1 - e^{-4\sigma_\mu \beta}} e^{-2\sigma_\mu \beta} < 3e^{-2\sigma_\mu \beta}, \\ |\zeta_n| &= \left| \frac{(1 - Q_n^2)e^{i\mu_n H}}{1 - Q_n^2 e^{2i\mu_n H}} \right| \le \frac{1 + e^{-4\sigma_\mu \beta}}{1 - e^{-4\sigma_\mu \beta}} < 1 + 3e^{-4\sigma_\mu \beta}, \end{aligned}$$

which yields the required inequalities (3.13). □

Lemma 3.3 *Let u be the solution to the problem* (3.8)-(3.9) *with $\gamma_{\text{in}}^{\text{L}} \in H^{-1/2}(\Gamma_{\text{L}})$ and $\gamma_{\text{in}}^{\text{R}} \in H^{-1/2}(\Gamma_{\text{R}})$. Then for σ_μ and β such that $e^{-2\sigma_\mu \beta} < 1/2$, the outgoing data defined by* (3.10) *satisfy*

$$\begin{aligned} \|\gamma_{\text{out}}^{\text{L}}\|_{H^{-1/2}(\Gamma_{\text{L}})} &\le 3e^{-2\sigma_\mu \beta} \|\gamma_{\text{in}}^{\text{L}}\|_{H^{-1/2}(\Gamma_{\text{L}})} + (1 + 3e^{-4\sigma_\mu \beta}) \|\gamma_{\text{in}}^{\text{R}}\|_{H^{-1/2}(\Gamma_{\text{R}})}, \\ \|\gamma_{\text{out}}^{\text{R}}\|_{H^{-1/2}(\Gamma_{\text{R}})} &\le (1 + 3e^{-4\sigma_\mu \beta}) \|\gamma_{\text{in}}^{\text{L}}\|_{H^{-1/2}(\Gamma_{\text{L}})} + 3e^{-2\sigma_\mu \beta} \|\gamma_{\text{in}}^{\text{R}}\|_{H^{-1/2}(\Gamma_{\text{R}})}. \end{aligned} \tag{3.14}$$

Proof Invoking (3.12) and Lemma 3.2, we can show that

$$\begin{aligned} |\gamma_{\text{out},n}^{\text{L}}| &\le 3e^{-2\sigma_\mu \beta} |\gamma_{\text{in},n}^{\text{L}}| + (1 + 3e^{-4\sigma_\mu \beta}) |\gamma_{\text{in},n}^{\text{R}}|, \\ |\gamma_{\text{out},n}^{\text{R}}| &\le (1 + 3e^{-4\sigma_\mu \beta}) |\gamma_{\text{in},n}^{\text{L}}| + 3e^{-2\sigma_\mu \beta} |\gamma_{\text{in},n}^{\text{R}}|. \end{aligned}$$

By using the triangle inequality of the norm in $H^{-1/2}(\Gamma)$ we are led to the estimates (3.14) for $\gamma_{\text{out}}^{\text{L}}$ and $\gamma_{\text{out}}^{\text{R}}$. □

3.3 Convergence of the continuous double sweep DDM

As seen in Remark 3.1 (3) the forward sweep process can be viewed as an operator $\mathbb{F} : G^{\text{R}} \to G^{\text{L}}$ defined by $\mathbb{F}(\boldsymbol{\gamma}^{\text{R},m}) = \boldsymbol{\gamma}^{\text{L},m}$, and the backward sweep process can be interpreted as an operator $\mathbb{B} : G^{\text{L}} \to G^{\text{R}}$ defined by $\mathbb{B}(\boldsymbol{\gamma}^{\text{L},m}) = \boldsymbol{\gamma}^{\text{R},m+1}$. Therefore, the double sweep process can be thought of a linear operator from the right boundary data $\boldsymbol{\gamma}^{\text{R},m}$ of the m-th iteration to the right boundary data $\boldsymbol{\gamma}^{\text{R},m+1}$ of the $(m+1)$-th iteration. For convergence of the error functions $\boldsymbol{u}^m - \boldsymbol{u}^{ex}$, assuming that $f_j = 0$ by linearity of the problem, we have only to show that $\mathbb{B} \circ \mathbb{F}$ is a contraction in G^{R}.

We first estimate the forward sweep operator $\mathbb{F}$ by using Lemma 3.3 repeatedly from Ω_1 to Ω_{J-1}.

Lemma 3.4 *Let $c_0 < 1/2$ be a positive constant such that $(6c_0^2 + 9c_0^4)(J-1) < 1$. If the* PML *parameters are chosen such that $e^{-2\sigma_\mu \beta} < c_0$, then the forward sweep operator $\mathbb{F}$ mapping from $\boldsymbol{\gamma}^{\mathrm{R},m}$ to $\boldsymbol{\gamma}^{\mathrm{L},m}$ satisfies the estimate*

$$\|\boldsymbol{\gamma}^{\mathrm{L},m}\|_{H^{-1/2}(\Gamma)} \lesssim (J-1)e^{-2\sigma_\mu \beta}\|\boldsymbol{\gamma}^{\mathrm{R},m}\|_{H^{-1/2}(\Gamma)}. \tag{3.15}$$

In particular,

$$\|\gamma_J^{\mathrm{L},m}\|_{H^{-1/2}(\Gamma_{J-1})} \lesssim \sqrt{J-1}e^{-2\sigma_\mu \beta}\|\boldsymbol{\gamma}^{\mathrm{R},m}\|_{H^{-1/2}(\Gamma)}. \tag{3.16}$$

Proof For the notational simplicity, we denote $\epsilon = e^{-2\sigma_\mu \beta}$. We begin with an estimation of $\gamma_2^{\mathrm{L},m}$ by applying Lemma 3.3 to the subdomain Ω_1 with $\gamma_1^{\mathrm{L},m} = 0$, which shows

$$\|\gamma_2^{\mathrm{L},m}\|_{H^{-1/2}(\Gamma_1)} \le 3\epsilon\|\gamma_1^{\mathrm{R},m}\|_{H^{-1/2}(\Gamma_1)}.$$

For $j = 3, 4, \ldots, J$, Lemma 3.3 leads us to

$$\|\gamma_j^{\mathrm{L},m}\|_{H^{-1/2}(\Gamma_{j-1})} \le 3\epsilon\|\gamma_{j-1}^{\mathrm{R},m}\|_{H^{-1/2}(\Gamma_{j-1})} + (1+3\epsilon^2)\|\gamma_{j-1}^{\mathrm{L},m}\|_{H^{-1/2}(\Gamma_{j-2})}.$$

The inductive argument for increasing j yields that

$$\|\gamma_j^{\mathrm{L},m}\|_{H^{-1/2}(\Gamma_{j-1})} \le 3\epsilon\sum_{i=1}^{j-1}(1+3\epsilon^2)^{j-i-1}\|\gamma_i^{\mathrm{R},m}\|_{H^{-1/2}(\Gamma_i)}. \tag{3.17}$$

Thus, each $\gamma_j^{\mathrm{L},m}$ for $j = 2, \ldots, J$ can be bounded by

$$\|\gamma_j^{\mathrm{L},m}\|^2_{H^{-1/2}(\Gamma_{j-1})} \le (3\epsilon)^2\sum_{i=1}^{J-1}(1+3\epsilon^2)^{2(i-1)}\|\boldsymbol{\gamma}^{\mathrm{R},m}\|^2_{H^{-1/2}(\Gamma)} \tag{3.18}$$

due to the Cauchy-Schwarz inequality. Noting that $(1+3\epsilon^2)^{2(J-1)} < e^{(6\epsilon^2+9\epsilon^4)(J-1)}$, one can easily show that for $(6\epsilon^2 + 9\epsilon^4)(J-1) < 1$

$$\sum_{i=1}^{J-1}(1+3\epsilon^2)^{2(i-1)} \le \frac{e^{(6\epsilon^2+9\epsilon^4)(J-1)} - 1}{6\epsilon^2 + 9\epsilon^4} < 2(J-1). \tag{3.19}$$

Combining (3.18) and (3.19) shows that for $j = 2, \ldots, J$

$$\|\gamma_j^{\mathrm{L},m}\|^2_{H^{-1/2}(\Gamma_{J-1})} \lesssim \epsilon^2(J-1)\|\boldsymbol{\gamma}^{\mathrm{R},m}\|^2_{H^{-1/2}(\Gamma)}. \tag{3.20}$$

Adding up (3.20) for $j = 2, \ldots, J$ gives us (3.15). Also, (3.16) is the result from (3.20) with $j = J$, which completes the proof. □

If we denote the forward sweep operator for the n-th mode by $\mathbb{F}_n$, then we find that $\mathbb{F}_n$ is a lower triangular Toeplitz matrix by using (3.12) repeatedly from Ω_1 to Ω_{J-1}, i.e., $\mathbb{F}_n = \varepsilon_n \Xi_n$, where

$$\Xi_n = \begin{bmatrix} 1 & & & & \\ \zeta_n & 1 & & & \\ \zeta_n^2 & \zeta_n & 1 & & \\ \vdots & \vdots & \ddots & \ddots & \\ \zeta_n^{J-2} & \zeta_n^{J-3} & \cdots & \zeta_n & 1 \end{bmatrix}. \tag{3.21}$$

Since $|\zeta_n| \approx 1$ for propagating modes, the norm of Ξ_n is of order $O(J-1)$ and hence the dependence of the stability constant on J appears to be inevitable.

The backward sweep operator $\mathbb{B}$ can be estimated in the similar fashion as done for the forward sweep operator $\mathbb{F}$ but with only a special care for the cavity Ω_J.

Lemma 3.5 *Let $c_0 < 1/2$ be a positive constant such that $(6c_0^2 + 9c_0^4)(J-1) < 1$. If the* PML *parameters are chosen such that $e^{-2\sigma_\mu \beta} < c_0$, then the backward sweep operator $\mathbb{B}$ mapping from $\boldsymbol{\gamma}^{\mathrm{L},m}$ to $\boldsymbol{\gamma}^{\mathrm{R},m+1}$ satisfies the estimate*

$$\|\boldsymbol{\gamma}^{\mathrm{R},m+1}\|_{H^{-1/2}(\Gamma)} \lesssim (J-1)e^{-2\sigma_\mu\beta}\|\boldsymbol{\gamma}^{\mathrm{L},m}\|_{H^{-1/2}(\Gamma)} + \sqrt{J-1}\|\gamma_J^{\mathrm{L},m}\|_{H^{-1/2}(\Gamma_{J-1})}. \tag{3.22}$$

Proof Let $\epsilon = e^{-2\sigma_\mu\beta}$. We first note that the solution $u_J^m \in V_J$ to the problem in Ω_J satisfies

$$\|u_J^m\|_{H^1(\Omega_J)} \lesssim \|\gamma_J^{\mathrm{L},m}\|_{H^{-1/2}(\Gamma_{J-1})}.$$

By a standard trace estimate with the stability of the solution we have

$$\|\gamma_{J-1}^{\mathrm{R},m+1}\|_{H^{-1/2}(\Gamma_{J-1})} = \|\frac{\partial u_J^m}{\partial \nu_J} + T_{\mathrm{PML}}(u_J^m)\|_{H^{-1/2}(\Gamma_{J-1})} \lesssim \|\gamma_J^{\mathrm{L},m}\|_{H^{-1/2}(\Gamma_{J-1})}. \tag{3.23}$$

Noting that

$$\|\gamma_j^{\mathrm{R},m+1}\|_{H^{-1/2}(\Gamma_j)} \le 3\epsilon\|\gamma_{j+1}^{\mathrm{L},m}\|_{H^{-1/2}(\Gamma_j)} + (1+3\epsilon^2)\|\gamma_{j+1}^{\mathrm{R},m+1}\|_{H^{-1/2}(\Gamma_{j+1})}$$

resulting from Lemma 3.3 applied to the subdomain Ω_{j+1} for $j = 1, \ldots, J-2$, the inductive argument for decreasing j from $j = J-2$ to $j = 1$ shows

$$\begin{aligned}\|\gamma_j^{\mathrm{R},m+1}\|_{H^{-1/2}(\Gamma_j)} \le 3\epsilon \left(\sum_{i=j+1}^{J-1} (1+3\epsilon^2)^{i-j-1}\|\gamma_i^{\mathrm{L},m}\|_{H^{-1/2}(\Gamma_{i-1})} \right) \\ + (1+3\epsilon^2)^{J-j-1}\|\gamma_{J-1}^{\mathrm{R},m+1}\|_{H^{-1/2}(\Gamma_{J-1})}.\end{aligned}$$

By the Cauchy-Schwarz inequality we can have

$$\begin{aligned}\|\gamma_j^{\mathrm{R},m+1}\|^2_{H^{-1/2}(\Gamma_j)} \le C\Bigg((3\epsilon)^2 \sum_{i=j+1}^{J-1} (1+3\epsilon^2)^{2(i-j-1)}\|\boldsymbol{\gamma}^{\mathrm{L},m}\|^2_{H^{-1/2}(\Gamma)} \\ + (1+3\epsilon^2)^{2(J-j-1)}\|\gamma_{J-1}^{\mathrm{R},m+1}\|^2_{H^{-1/2}(\Gamma_{J-1})}\Bigg).\end{aligned} \tag{3.24}$$

Since the condition $(6\epsilon^2 + 9\epsilon^4)(J-1) < 1$ implies

$$(1+3\epsilon^2)^{2(J-j-1)} < (1+6\epsilon^2+9\epsilon^4)^{J-1} < e^{(6\epsilon^2+9\epsilon^4)(J-1)} < e,$$

invoking the same estimate as (3.19) in the forward sweep, we can show that

$$\|\gamma_j^{\mathrm{R},m+1}\|^2_{H^{-1/2}(\Gamma_j)} \lesssim \epsilon^2(J-1)\|\boldsymbol{\gamma}^{\mathrm{L},m}\|^2_{H^{-1/2}(\Gamma)} + \|\gamma_{J-1}^{\mathrm{R},m+1}\|^2_{H^{-1/2}(\Gamma_{J-1})} \tag{3.25}$$

Finally, (3.22) results from adding all (3.25) for $j = 1, \ldots, J-1$ and then applying (3.23). □

Now, we are in a position to show that the double sweep operator $\mathbb{B} \circ \mathbb{F}$ is a contraction for sufficiently small $e^{-2\sigma_\mu \beta}$.

Lemma 3.6 *Let c_0 be a positive constant such that $c_0(J-1) < 1$ as well as $(6c_0^2 + 9c_0^4)(J-1) < 1$. If the* PML *parameters are chosen such that $e^{-2\sigma_\mu \beta} < c_0$, then the double sweep operator $\mathbb{B} \circ \mathbb{F}$ mapping from $\boldsymbol{\gamma}^{\mathrm{R},m}$ to $\boldsymbol{\gamma}^{\mathrm{R},m+1}$ satisfies the estimate*

$$\|\boldsymbol{\gamma}^{\mathrm{R},m+1}\|_{H^{-1/2}(\Gamma)} \leq C_{\mathbb{B}\circ\mathbb{F}}(J-1)e^{-2\sigma_\mu \beta}\|\boldsymbol{\gamma}^{\mathrm{R},m}\|_{H^{-1/2}(\Gamma)}$$

for some positive constant $C_{\mathbb{B}\circ\mathbb{F}}$ that may depend only on k and Ω_J.

Proof Let $\epsilon = e^{-2\sigma_\mu \beta}$. We use Lemma 3.4 and Lemma 3.5 to show that

$$\begin{aligned}\|\boldsymbol{\gamma}^{\mathrm{R},m+1}\|^2_{H^{-1/2}(\Gamma)} &\lesssim \epsilon^2(J-1)^2\|\boldsymbol{\gamma}^{\mathrm{L},m}\|^2_{H^{-1/2}(\Gamma)} + (J-1)\|\gamma_J^{\mathrm{L},m}\|^2_{H^{-1/2}(\Gamma_{J-1})} \\ &\lesssim (\epsilon^2(J-1)^2 + 1)\epsilon^2(J-1)^2\|\boldsymbol{\gamma}^{\mathrm{R},m}\|^2_{H^{-1/2}(\Gamma)}.\end{aligned}$$

Since $\epsilon^2(J-1)^2 < 1$, the desired estimate is achieved. □

Remark 3.7 *If Ω_J is also open to the right, then the constant $C_{\mathbb{B}\circ\mathbb{F}}$ becomes a generic constant independent of k and Ω_J according to the proof of Lemma 3.5 and Lemma 3.6.*

Theorem 3.8 *Suppose that $f = 0$. Let $\boldsymbol{u}^m \in \boldsymbol{V}$ be the m-th iterate of the double sweep DDM for any $\boldsymbol{u}^0$ satisfying* (3.7)*. Assume that $c_0 < 1/2$ is a positive constant satisfying $c_0(J-1) < 1$ as well as $(6c_0^2 + 9c_0^4)(J-1) < 1$. If the* PML *parameters are chosen such that $e^{-2\sigma_\mu \beta} < c_0$, then it holds that*

$$\|\boldsymbol{u}^m\|_{\boldsymbol{V}} \lesssim (C_{\mathbb{B}\circ\mathbb{F}}(J-1)e^{-2\sigma_\mu \beta})^m\|\boldsymbol{u}^0\|_{\boldsymbol{V}}.$$

Proof Since the m-th iterate $\boldsymbol{u}^m$ is determined by $\boldsymbol{\gamma}^{\mathrm{R},m}$ and $\boldsymbol{\gamma}^{\mathrm{L},m-1}$, by the stability of local problems we have

$$\|\boldsymbol{u}^m\|_{\boldsymbol{V}} \lesssim \|\boldsymbol{\gamma}^{\mathrm{R},m}\|_{H^{-1/2}(\Gamma)} + \|\boldsymbol{\gamma}^{\mathrm{L},m-1}\|_{H^{-1/2}(\Gamma)},$$

which can be in turn by Lemma 3.4 and Lemma 3.6

$$\|\boldsymbol{u}^m\|_{\boldsymbol{V}} \lesssim C_{\mathbb{B}\circ\mathbb{F}}(J-1)e^{-2\sigma_\mu \beta}\|\boldsymbol{\gamma}^{\mathrm{R},m-1}\|_{H^{-1/2}(\Gamma)}.$$

Finally, Lemma 3.6 and a trace inequality reveal that

$$\begin{aligned}\|\boldsymbol{u}^m\|_{\boldsymbol{V}} &\lesssim (C_{\mathbb{B}\circ\mathbb{F}}(J-1)e^{-2\sigma_\mu \beta})^m\|\boldsymbol{\gamma}^{\mathrm{R},0}\|_{H^{-1/2}(\Gamma)} \\ &\lesssim (C_{\mathbb{B}\circ\mathbb{F}}(J-1)e^{-2\sigma_\mu \beta})^m\|\boldsymbol{u}^0\|_{\boldsymbol{V}},\end{aligned}$$

which completes the proof. □

As a consequence, the double sweep DDM converges if we apply PML such that $C_{\mathbb{B}\circ\mathbb{F}}(J-1)e^{-2\sigma_\mu \beta} < 1$.

4 Analysis of the finite element subdomain transmission problem

In this section we study the finite element transmission problem on the reference domain $\hat{\Omega}$ that will be used to analyze the convergence of the double sweep DDM applied to the Helmholtz equation discretized by the finite element method.

We first introduce the finite element spaces on the whole domain $\widetilde{\Omega}$ and on the reference domain $\hat{\Omega}$. Let $h = H/N_s = \beta/N_p$ for positive integers N_s and N_p. We define a quasi-uniform and shape-regular interval/quadrilateral mesh $\mathcal{T}_\Gamma$ of maximal diameter h on Γ and identify the triangulation $\mathcal{T}_{\Gamma_j}$ on Γ_j with $\mathcal{T}_\Gamma$. We introduce quasi-uniform and shape-regular quadrilateral/hexahedral meshes $\mathcal{T}_{\Omega_j}$ and $\mathcal{T}_{\Omega_{\mathrm{PML}}}$ of subdomains Ω_j and Ω_{PML}, respectively, which is obtained by starting with the triangulation of $\mathcal{T}_{\Gamma_j}$ and extruding elements in $\mathcal{T}_{\Gamma_j}$ by every h along the axis of the waveguide. Thus, the common interface Γ_j of Ω_j and Ω_{j+1} for $j = 1, \ldots, J-1$ has the same face triangulation $\mathcal{T}_{\Gamma_j}$ inherited from Ω_j and Ω_{j+1}. Let $\mathcal{T}_\Omega = \cup_{j=1}^J \mathcal{T}_{\Omega_j}$ for the finite element mesh of Ω and $\mathcal{T}_{\widetilde{\Omega}} = \mathcal{T}_\Omega \cup \mathcal{T}_{\Omega_{\mathrm{PML}}}$ for that of $\widetilde{\Omega}$. By $V^h_{\widetilde{\Omega}}$, V^h_Ω, $V^h_{\Omega_j}$ and $V^h_{\Omega_{\mathrm{PML}}}$ we denote finite element spaces of continuous piecewise bilinear/trilinear functions defined in $\widetilde{\Omega}$, Ω, Ω_j and Ω_{PML} corresponding to their triangulations, respectively. $V^h_{\Gamma_j}$ is the finite element space induced by the trace of functions in $V^h_{\Omega_j}$ and V^h_Γ can be defined by its identification with $V^h_{\Gamma_j}$. We can also define $\mathcal{T}_{\hat{\Omega}}$, $\mathcal{T}_{\Gamma_{\mathrm{L}}}$, $\mathcal{T}_{\Gamma_{\mathrm{R}}}$ and $V^h_{\hat{\Omega}}$, $V^h_{\Gamma_{\mathrm{L}}}$, $V^h_{\Gamma_{\mathrm{R}}}$, analogously.

4.1 Solution representation in $\hat{\Omega}$

The interval (L, R) is decomposed into the N_s numbers of uniform subintervals

$$\mathrm{L} = a_0 < a_1 < \ldots < a_{N_s} = \mathrm{R}, \quad a_q - a_{q-1} = \frac{H}{N_s} = h \text{ for } q = 1, \ldots, N_s.$$

Then the triangulations at $\{a_q\} \times \Gamma$ are all identical for $q = 0, \ldots, N_s$. We will use $\mathcal{V}^h_{\Gamma,q}$ for the finite element spaces of continuous piecewise bilinear/trilinear finite element functions vanishing on all nodes outside of $\{a_q\} \times \Gamma$ as a subspace of $V^h_{\hat{\Omega}}$. Let N_Γ denote the dimension of $\mathcal{V}^h_{\Gamma,q}$. Note that $\mathcal{V}^h_{\Gamma,q}$ is different from the finite element space $V^h_{\Gamma_j}$ on the interface Γ_j, although they have the same dimension. We see that every function u_h in the finite element space $V^h_{\hat{\Omega}}$ of dimension $N_\Gamma \times (N_s + 1)$ can be written uniquely as $u_h = \sum_{q=0}^{N_s} u_{h,q}$ with $u_{h,q} \in \mathcal{V}^h_{\Gamma,q}$. With the same ordering of the nodal finite element basis functions in every $\mathcal{V}^h_{\Gamma,q}$, we denote by $\hat{u}_{h,q}$ the coordinate representation of $u_{h,q}$ with respect to the nodal finite element basis. Then $\hat{u}_h = (\hat{u}_{h,0}, \ldots, \hat{u}_{h,N_s})$ is the corresponding vector in $\mathbb{C}^{N_\Gamma \times (N_s+1)}$ for u_h.

Let us first consider the finite element approximation to the eigenvalue problem (2.2) on the cross-section Γ. Denoting the finite element approximations to the eigenvalue problem (2.2) by $(\phi_{h,\ell}, \lambda^2_{h,\ell}) \in V^h_\Gamma \times \mathbb{R}$ for $\ell = 1, 2, \ldots, N_\Gamma$ with $N_\Gamma = \dim(V^h_\Gamma)$, the coordinate representations $\hat{\phi}_{h,\ell} \in \mathbb{R}^{N_\Gamma}$ of $\phi_{h,\ell}$ with respect to the nodal finite element basis functions are solutions to the eigenvalue problem in a finite dimensional space

$$S_\Gamma \hat{\phi}_{h,\ell} = \lambda^2_{h,\ell} M_\Gamma \hat{\phi}_{h,\ell}, \tag{4.1}$$

where S_Γ and M_Γ are the $N_\Gamma \times N_\Gamma$ stiffness and mass matrices, respectively. We assume that eigenvalues are ordered increasingly,

$$\lambda_{h,1} \le \lambda_{h,2} \le \ldots \le \lambda_{h,N_\Gamma}.$$

Here we note that $\lambda_{h,\ell} \ge 0$ and the largest eigenvalue $\lambda^2_{h,N_\Gamma} = O(h^{-2})$ by the inverse inequality so that there exists a positive constant C_λ such that $\lambda^2_{h,\ell} h^2 \le C^2_\lambda$. Also, we can choose the eigenvectors $\hat{\phi}_{h,\ell}$ of (4.1), which are orthonormal with respect to the M_Γ-inner product defined by $(x,y)_{M_\Gamma} := (M_\Gamma x, y)$ for $x, y \in \mathbb{C}^{N_\Gamma}$. Due to the interpolation theory [1,4,31], it can be shown that the discrete fractional order norm defined by

$$\|w_h\|^2_{\dot{H}^s_h(\Gamma)} := \sum_{\ell=1}^{N_\Gamma} (1+\lambda^2_{h,\ell})^s |w^h_\ell|^2$$

for $w_h \in V^h_\Gamma$ and $-1 \le s \le 1$, is equivalent to the continuous fractional order norm, where $w^h_\ell = (w_h, \phi_{h,\ell})_\Gamma$, that is, there exist positive constants C_1 and C_2 independent of h such that

$$C_1 \|w_h\|_{H^s(\Gamma)} \le \|w_h\|_{\dot{H}^s_h(\Gamma)} \le C_2 \|w_h\|_{H^s(\Gamma)} \tag{4.2}$$

for $w_h \in V^h_\Gamma$.

Now, for given $\gamma^{\mathrm{L}}_h \in V^h_{\Gamma_{\mathrm{L}}}$ and $\gamma^{\mathrm{R}}_h \in V^h_{\Gamma_{\mathrm{R}}}$ consider the problem to find $u_h \in V^h_{\hat{\Omega}}$ such that

$$b^h_{\hat{\Omega}}(u_h, v_h) = (\gamma^{\mathrm{L}}_h, v_h)_{\Gamma_{\mathrm{L}}} + (\gamma^{\mathrm{R}}_h, v_h)_{\Gamma_{\mathrm{R}}} \text{ for all } v_h \in V^h_{\hat{\Omega}}, \tag{4.3}$$

where

$$b^h_{\hat{\Omega}}(u_h, v_h) = a_{\hat{\Omega}}(u_h, v_h) - (T^h_{\mathrm{PML}}(u_h), v_h)_{\Gamma_{\mathrm{L}} \cup \Gamma_{\mathrm{R}}}. \tag{4.4}$$

with

$$a_{\hat{\Omega}}(u_h, v_h) = (\nabla u_h, \nabla v_h)_{\hat{\Omega}} - k^2 (u_h, v_h)_{\hat{\Omega}}. \tag{4.5}$$

Here T^h_{PML} is the discrete PML operator, that will be discussed in Subsection 4.3. We will consider $b^h_j(\cdot,\cdot) = b^h_{\hat{\Omega}}(\cdot,\cdot)$ and $a_j(\cdot,\cdot) = a_{\hat{\Omega}}(\cdot,\cdot)$ with $\hat{\Omega} = \Omega_j$ in dealing with the local subdomain problems later. In this case, the system matrix for the sesquilinear form $b^h_{\hat{\Omega}}(\cdot,\cdot)$ can be written as the block tridiagonal matrix

$$\begin{bmatrix} A - \hat{T}^h_{\mathrm{PML}} & -B & & & \\ -B & 2A & -B & & \\ & \ddots & \ddots & \ddots & \\ & & -B & 2A & -B \\ & & & -B & A - \hat{T}^h_{\mathrm{PML}} \end{bmatrix} \tag{4.6}$$

where A in the diagonal blocks is an $N_\Gamma \times N_\Gamma$ matrix related with interaction of two basis functions supported on the same cross-section whereas B in the off-diagonal blocks is one that results from two basis functions supported on two different neighboring cross-sections. $\hat{T}^h_{\mathrm{PML}}$ stands for the matrix corresponding to the discrete PML operator T^h_{PML}, which also will be studied in Subsection 4.3 together with T^h_{PML}. By splitting the matrices A and B into three parts corresponding to

each term in (4.5), $A = A_x + A_y - k^2 A_0$ and $B = B_x + B_y - k^2 B_0$, respectively, we obtain that

$$A = \frac{h}{3} S_\Gamma + (\frac{1}{h} - \frac{hk^2}{3}) M_\Gamma, \qquad B = \frac{-h}{6} S_\Gamma + (\frac{1}{h} + \frac{hk^2}{6}) M_\Gamma. \tag{4.7}$$

Indeed, noting that bilinear/trilinear functions are obtained by tensor products, it can be shown that

$$\begin{aligned} A_x &= \frac{1}{h} M_\Gamma, \qquad & A_y &= \frac{h}{3} S_\Gamma, \qquad & A_0 &= \frac{h}{3} M_\Gamma, \\ B_x &= \frac{1}{h} M_\Gamma, \qquad & B_y &= \frac{-h}{6} S_\Gamma, \qquad & B_0 &= \frac{-h}{6} M_\Gamma. \end{aligned} \tag{4.8}$$

Assuming that the mesh size h satisfies $h < 1/(\varrho k)$ with mesh density $\varrho \geq 2$ so that there exist a certain amount of grid points per wavelength for resolving wave phenomena properly, the positivity of $\lambda_{h,\ell}^2 \geq 0$ yields that

$$h^2 \mu_{h,\ell}^2 \leq h^2 k^2 \leq \frac{1}{\varrho^2}, \tag{4.9}$$

where $\mu_{h,\ell} = \sqrt{k^2 - \lambda_{h,\ell}^2}$.

Now, the invertibility of the matrices A and B is discussed in the following lemma. To do this, let J_B be a set of indices ℓ such that $1 + h^2 \mu_{h,\ell}^2 / 6 = 0$.

Lemma 4.1 *The matrix A is invertible. If $J_B = \varnothing$, then the matrix B is invertible as well. If $J_B \neq \varnothing$, then eigenvectors $\hat{\phi}_{h,\ell}$ for $\ell \in J_B$ generate the null space of B.*

Proof The formula (4.7) implies that

$$A\hat{\phi}_{h,\ell} = \frac{1}{h} \left(1 - \frac{h^2 \mu_{h,\ell}^2}{3} \right) M_\Gamma \hat{\phi}_{h,\ell}, \tag{4.10}$$

$$B\hat{\phi}_{h,\ell} = \frac{1}{h} \left(1 + \frac{h^2 \mu_{h,\ell}^2}{6} \right) M_\Gamma \hat{\phi}_{h,\ell}. \tag{4.11}$$

Noting that $\{\hat{\phi}_{h,\ell}\}_{\ell=1}^{N_\Gamma}$ forms a basis in $\mathbb{C}^{N_\Gamma}$, since $1 - h^2 \mu_{h,\ell}^2 / 3 \neq 0$ due to (4.9) and M_Γ is invertible, the matrix A maps a basis to another basis, which implies that A is invertible. Similarly, if $J_B = \varnothing$, by the same argument as above, B is invertible. When $J_B \neq \varnothing$, (4.11) yields that $B\hat{\phi}_{h,\ell} = 0$ for $\ell \in J_B$ and the vectors $B\hat{\phi}_{h,\ell}$ for $\ell \notin J_B$ are linearly independent, which completes the proof. □

Lemma 4.2 *If $\ell \notin J_B$, then $\hat{\phi}_{h,\ell}$ is an eigenvector of the generalized eigenvalue problem*

$$A\hat{\phi} = \eta B\hat{\phi} \tag{4.12}$$

for an eigenvalue

$$\eta_{h,\ell} = \left(1 - \frac{h^2 \mu_{h,\ell}^2}{3} \right) \left(1 + \frac{h^2 \mu_{h,\ell}^2}{6} \right)^{-1}.$$

Proof It follows from (4.10) and (4.11). □

In order to study solutions to the problem (4.3), which satisfy three term recurrence

$$-B\hat{u}_{h,q-1} + 2A\hat{u}_{h,q} - B\hat{u}_{h,q+1} = 0 \text{ for } q = 1, 2, \ldots, N_s - 1, \tag{4.13}$$

we examine solutions to the characteristic equation $\xi^2 - 2\eta_{h,\ell}\xi + 1 = 0$ for each $\eta_{h,\ell}$, $\ell \notin J_B$. The eigenvalue $\eta_{h,\ell}$ is real and its absolute value is classified into three possible cases:

1 For $|\eta_{h,\ell}| = 1$, the characteristic equation has a multiple root, however this can not occur for sufficiently small $h > 0$. Indeed, $|\eta_{h,\ell}| = 1$ if and only if $h^2\mu_{h,\ell}^2 = 0$ or 12. The condition (4.9) for the mesh resolution gives $h^2\mu_{h,\ell}^2 \neq 12$. In addition, the assumption excluding cutoff modes with sufficiently small h guarantees that $\mu_{h,\ell}^2 = k^2 - \lambda_{h,\ell}^2 \neq 0$, which results in $h^2\mu_{h,\ell}^2 \neq 0$.

2 $|\eta_{h,\ell}| < 1$ if and only if $0 < h^2\mu_{h,\ell}^2 < 12$. By the assumption (4.9) saying that $h^2\mu_{h,\ell}^2 < 1/\varrho^2 < 1$ for all $\ell = 1, 2, \ldots, N_\Gamma$, we have only $0 < \eta_{h,\ell} < 1$ and in this case the solution $\xi_{h,\ell}$ with $\Im(\xi_{h,\ell}) > 0$ of the characteristic equation can be written as

$$\xi_{h,\ell} = \eta_{h,\ell} + i\sqrt{1 - \eta_{h,\ell}^2} = e^{ih\mu_{h,\ell}^*} \quad \text{if } 0 < h^2\mu_{h,\ell}^2 < 1/\varrho^2 \ (0 < \eta_{h,\ell} < 1) \tag{4.14}$$

for some $\mu_{h,\ell}^* > 0$. We note that the condition $\mu_{h,\ell}^2 > 0$ corresponds to that for propagating modes of the continuous problem and in fact, these modes represent discrete propagating modes approximating continuous propagating modes. For such $\mu_{h,\ell}$ with $0 < h\mu_{h,\ell} < 1/\varrho$, $\eta_{h,\ell}$ is a Padé approximation to $\cos(h\mu_{h,\ell})$,

$$\eta_{h,\ell} = \left(1 - \frac{h^2\mu_{h,\ell}^2}{3}\right)\left(1 + \frac{h^2\mu_{h,\ell}^2}{6}\right)^{-1} = 1 - \frac{h^2\mu_{h,\ell}^2}{2} + O((h\mu_{h,\ell})^4)$$

and hence $\xi_{h,\ell}$ is an approximation of $e^{i\mu_n h}$ for a propagation mode at $x = h$,

$$\xi_{h,\ell} = e^{ih\mu_{h,\ell}^*} \approx e^{ih\mu_{h,\ell}} \approx e^{ih\mu_n}$$

for some $\mu_n > 0$, a wavenumber of a propagating mode. Thus, we call $\mu_{h,\ell}$ and $\mu_{h,\ell}^*$ the discrete wavenumber and the numerical wavenumber of the ℓ-th mode, respectively, corresponding to the continuous wavenumber μ_n of the n-th propagating mode. It is known that the discrepancy between $\mu_{h,\ell}^*$ and μ_n is the main source for the pollution error in finite element approximations for wave propagation problems [22, 23, 24].

3 $|\eta_{h,\ell}| > 1$ if and only if $h^2\mu_{h,\ell}^2 < 0$ or $h^2\mu_{h,\ell}^2 > 12$. By the assumption on the mesh resolution, we have only $h^2\mu_{h,\ell}^2 < 0$. It is equivalent to the condition that $\mu_{h,\ell}^2 < 0$, which is the same condition as that for evanescent modes of the continuous problem. In this case, we choose

$$\xi_{h,\ell} = \begin{cases} \eta_{h,\ell} - \sqrt{\eta_{h,\ell}^2 - 1} & \text{if } \ -6 < h^2\mu_{h,\ell}^2 < 0 \ \ (\eta_{h,\ell} > 1), \\ \eta_{h,\ell} + \sqrt{\eta_{h,\ell}^2 - 1} & \text{if } \ h^2\mu_{h,\ell}^2 < -6 \ \ (\eta_{h,\ell} < -2) \end{cases} \tag{4.15}$$

satisfying $|\xi_{h,\ell}| < 1$. See the plot of $|\xi|$ as a function of $x = (h\mu)^2 < 0$ in Fig. 2.

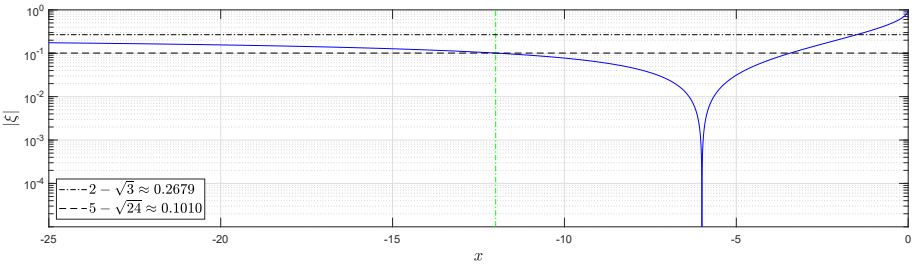


Fig. 2: Plot of $|\xi|$ as a function of $x = (h\mu)^2 < 0$ for evanescent modes. Here $\lim_{x\to-\infty}|\xi(x)| = 2-\sqrt{3}$ and $|\xi(-12)| = 5-\sqrt{24}$. The green vertical line represents the lower bound -12 of x in two dimensional problem as the cross-sectional discrete Neumann eigenvalues for $\Gamma = (0,1)$ are $\lambda_{h,\ell}^2 = 12h^{-2}\sin^2(\frac{\ell\pi}{2N_\Gamma})/(1+2\cos^2(\frac{\ell\pi}{2N_\Gamma}))$ and hence $(h\mu_{h,N_\Gamma})^2 = h^2k^2 - h^2\lambda_{h,N_\Gamma}^2 \geq -12$.

Lemma 4.3 *Suppose that* $-B\hat{u}_{h,q-1}+2A\hat{u}_{h,q}-B\hat{u}_{h,q+1} = 0$ *for* $q = 1,2,\ldots,N_s-1$. *Then* $\hat{u}_{h,q}$ *for* $q = 0,1,\ldots,N_s$ *is of the form*

$$\hat{u}_{h,q} = \sum_{\ell\notin J_B}(\alpha_\ell\xi_{h,\ell}^q + \beta_\ell\xi_{h,\ell}^{-q})\hat{\phi}_{h,\ell} + (\delta_{0,q}+\delta_{N_s,q})\sum_{\ell\in J_B} r_{q,\ell}\hat{\phi}_{h,\ell} \tag{4.16}$$

for constants α_ℓ, β_ℓ *and* $r_{0,\ell}, r_{N_s,\ell}$, *where* $\delta_{\ell,q}$ *is the Kronecker delta and* $\xi_{h,\ell}$ *is the solution to the characteristic equation* $\xi^2 - 2\eta_{h,\ell}\xi + 1 = 0$ *given by* (4.14) *or* (4.15) *depending on* $\eta_{h,\ell}$.

Proof We first consider the case that $J_B = \varnothing$, that is, B is invertible. In this case, $\hat{u}_{h,q}$ satisfies

$$L\begin{bmatrix}\hat{u}_{h,q}\\ \hat{u}_{h,q-1}\end{bmatrix} := \begin{bmatrix}2B^{-1}A & -I\\ I & 0\end{bmatrix}\begin{bmatrix}\hat{u}_{h,q}\\ \hat{u}_{h,q-1}\end{bmatrix} = \begin{bmatrix}\hat{u}_{h,q+1}\\ \hat{u}_{h,q}\end{bmatrix},$$

from which we see that

$$\begin{bmatrix}\hat{u}_{h,q+1}\\ \hat{u}_{h,q}\end{bmatrix} = L^q\begin{bmatrix}\hat{u}_{h,1}\\ \hat{u}_{h,0}\end{bmatrix} \text{ for } q = 0,1,\ldots,N_s-1. \tag{4.17}$$

Since it holds that

$$L\begin{bmatrix}\xi_{h,\ell}^{\pm1}\hat{\phi}_{h,\ell}\\ \hat{\phi}_{h,\ell}\end{bmatrix} = \begin{bmatrix}(2\eta_{h,\ell}\xi_{h,\ell}^{\pm1}-1)\hat{\phi}_{h,\ell}\\ \xi_{h,\ell}^{\pm1}\hat{\phi}_{h,\ell}\end{bmatrix} = \xi_{h,\ell}^{\pm1}\begin{bmatrix}\xi_{h,\ell}^{\pm1}\hat{\phi}_{h,\ell}\\ \hat{\phi}_{h,\ell}\end{bmatrix},$$

$\left\{\begin{bmatrix}\xi_{h,\ell}\hat{\phi}_{h,\ell}\\ \hat{\phi}_{h,\ell}\end{bmatrix}, \begin{bmatrix}\xi_{h,\ell}^{-1}\hat{\phi}_{h,\ell}\\ \hat{\phi}_{h,\ell}\end{bmatrix}\right\}_{\ell=1}^{N_\Gamma}$ is a basis consisting of eigenvectors of L for eigenvalues $\{\xi_{h,\ell}, \xi_{h,\ell}^{-1}\}_{\ell=1}^{N_\Gamma}$. Therefore, $\begin{bmatrix}\hat{u}_{h,1}\\ \hat{u}_{h,0}\end{bmatrix}$ can be written as a linear combination of the eigenvectors of L,

$$\begin{bmatrix}\hat{u}_{h,1}\\ \hat{u}_{h,0}\end{bmatrix} = \sum_{\ell=1}^{N_\Gamma}\alpha_\ell\begin{bmatrix}\xi_{h,\ell}\hat{\phi}_{h,\ell}\\ \hat{\phi}_{h,\ell}\end{bmatrix} + \beta_\ell\begin{bmatrix}\xi_{h,\ell}^{-1}\hat{\phi}_{h,\ell}\\ \hat{\phi}_{h,\ell}\end{bmatrix}.$$

Invoking (4.17), we obtain that

$$\begin{bmatrix} \hat{u}_{h,q+1} \\ \hat{u}_{h,q} \end{bmatrix} = \sum_{\ell=1}^{N_\Gamma} \alpha_\ell \xi_{h,\ell}^q \begin{bmatrix} \xi_{h,\ell}\hat{\phi}_{h,\ell} \\ \hat{\phi}_{h,\ell} \end{bmatrix} + \beta_\ell \xi_{h,\ell}^{-q} \begin{bmatrix} \xi_{h,\ell}^{-1}\hat{\phi}_{h,\ell} \\ \hat{\phi}_{h,\ell} \end{bmatrix}.$$

As a consequence, by taking the second component, (4.16) is established.

In case that $J_B \neq \varnothing$ (B is not invertible), we write

$$\hat{u}_{h,q} = \sum_{\ell=1}^{N_\Gamma} r_{q,\ell}\hat{\phi}_{h,\ell}$$

for $q = 0, 1, \ldots, N_s$, since $\{\hat{\phi}_{h,\ell}\}_{\ell=1}^{N_\Gamma}$ forms a basis. Noting that both A and B are bijective from the subspace spanned by $\{\hat{\phi}_{h,\ell}\}_{\ell\notin J_B}$ to the subspace spanned by $\{M_\Gamma\hat{\phi}_{h,\ell}\}_{\ell\notin J_B}$, the same argument used as above by applying to the subspace spanned by $\{\hat{\phi}_{h,\ell}\}_{\ell\notin J_B}$ can show that $r_{q,\ell} = \alpha_\ell\xi_{h,\ell}^q + \beta_\ell\xi_{h,\ell}^{-q}$ for $\ell \notin J_B$. Since the coefficients of $\hat{\phi}_{h,\ell}$ for $\ell \notin J_B$ in (4.13) vanishes, it follows that

$$0 = -B\hat{u}_{h,q-1} + 2A\hat{u}_{h,q} - B\hat{u}_{h,q+1} = 2\sum_{\ell\in J_B} r_{q,\ell}A\hat{\phi}_{h,\ell}.$$

$q = 1, 2, \ldots, N_s - 1$. By linear independence of $\{A\hat{\phi}_{h,\ell}\}_{\ell\in J_B}$, it can be concluded that $r_{q,\ell} = 0$ for $q = 1, 2, \ldots, N_s - 1$ and $\ell \in J_B$. In other words, coefficients $r_{q,\ell}$ may not vanish only if $\ell \in J_B$, and $q = 0$ or N_s, which completes the proof. □

4.2 Discrete normal derivatives

In this subsection, we define the variational discrete normal derivatives on Γ_{L} and Γ_{R} of finite element functions and derive their coordinate representations.

Let $u_h \in V_{\hat{\Omega}}^h$ be the solution to the problem (4.3), which has a decomposition $u_h = \sum_{q=0}^{N_s} u_{h,q}$ with $u_{h,q} \in \mathcal{V}_{\Gamma,q}^h$. Since u_h satisfies $a_{\hat{\Omega}}(u_h, v_h^0) = 0$ for all $v_h^0 \in V_{\hat{\Omega}}^h$ vanishing on Γ_{L} and Γ_{R}, the discrete normal derivative of u_h on Γ_{L} from $\hat{\Omega}$, denoted by $\frac{\partial^h u_h}{\partial\nu}$ in $V_{\Gamma_{\mathrm{L}}}^h$, is defined in the variational sense by

$$(\frac{\partial^h u_h}{\partial\nu}, v_h)_{\Gamma_{\mathrm{L}}} = a_{\hat{\Omega}}(u_h, \tilde{v}_h) \text{ for } v_h \in V_{\Gamma_{\mathrm{L}}}^h, \tag{4.18}$$

where $\tilde{v}_h \in V_{\hat{\Omega}}^h$ is any extension of v_h vanishing on Γ_{R}. The discrete normal derivative on Γ_{R} can be defined analogously.

Lemma 4.4 *Let $\hat{u}_h = (\hat{u}_{h,0}, \ldots, \hat{u}_{h,N_s})$ be the coordinate representation of u_h corresponding to the decomposition of $u_h = \sum_{q=0}^{N_s} u_{h,q}$, whose components are of the form* (4.16)*. Then the discrete normal derivative $\frac{\partial^h u_h}{\partial\nu} \in V_{\Gamma_{\mathrm{L}}}^h$ of u_h on Γ_{L} defined by* (4.18) *and its analogous one on Γ_{R} have the coordinate representations*

$$\frac{\partial^h \hat{u}_h}{\partial\nu} = \sum_{\ell\notin J_B} -\Lambda_{h,\ell}(\alpha_\ell - \beta_\ell)\hat{\phi}_{h,\ell} + \sum_{\ell\in J_B} \frac{3}{h} r_{0,\ell}\hat{\phi}_{h,\ell} \text{ on } \Gamma_{\mathrm{L}}, \tag{4.19}$$

and

$$\frac{\partial^h \hat{u}_h}{\partial \nu} = \sum_{\ell \notin J_B} \Lambda_{h,\ell}(\alpha_\ell \xi_{h,\ell}^{N_s} - \beta_\ell \xi_{h,\ell}^{-N_s})\hat{\phi}_{h,\ell} + \sum_{\ell \in J_B} \frac{3}{h} r_{N_s,\ell} \hat{\phi}_{h,\ell} \text{ on } \Gamma_{\mathrm{R}}, \tag{4.20}$$

respectively, where

$$\Lambda_{h,\ell} := i\mu_{h,\ell}\sqrt{1 - \frac{h^2\mu_{h,\ell}^2}{12}}.$$

Proof The discrete normal derivatives $\frac{\partial^h u_h}{\partial \nu}$ of u_h on Γ_{L} and on Γ_{R} have the coordinate representation, $\frac{\partial^h \hat{u}_h}{\partial \nu}$, satisfying

$$M_\Gamma \frac{\partial^h \hat{u}_h}{\partial \nu} = A\hat{u}_{h,0} - B\hat{u}_{h,1} \text{ on } \Gamma_{\mathrm{L}}, \tag{4.21}$$

$$M_\Gamma \frac{\partial^h \hat{u}_h}{\partial \nu} = -B\hat{u}_{h,N_s-1} + A\hat{u}_{h,N_s} \text{ on } \Gamma_{\mathrm{R}}, \tag{4.22}$$

respectively. From (4.21) and the solution representation (4.16), it can be shown that on Γ_{L}

$$\begin{aligned} M_\Gamma \frac{\partial^h \hat{u}_h}{\partial \nu} &= \sum_{\ell \notin J_B} \left[(\alpha_\ell + \beta_\ell)\eta_{h,\ell} - (\alpha_\ell \xi_{h,\ell} + \beta_\ell \xi_{h,\ell}^{-1})\right] B\hat{\phi}_{h,\ell} + \sum_{\ell \in J_B} r_{0,\ell} A\hat{\phi}_{h,\ell} \\ &= \sum_{\ell \notin J_B} \frac{1}{2}(\xi_{h,\ell}^{-1} - \xi_{h,\ell})(\alpha_\ell - \beta_\ell) B\hat{\phi}_{h,\ell} + \sum_{\ell \in J_B} \frac{3}{h} r_{0,\ell} M_\Gamma \hat{\phi}_{h,\ell}. \end{aligned} \tag{4.23}$$

Noting the choice of $\xi_{h,\ell}$, (4.14) and (4.15) for $\ell \notin J_B$, it holds that

$$\frac{1}{2}(\xi_{h,\ell}^{-1} - \xi_{h,\ell}) = \begin{cases} -\sqrt{\eta_{h,\ell}^2 - 1} & \text{if } h^2\mu_{h,\ell}^2 \in (-\infty, -6) \cup (0, 1/\varrho^2), \\ \sqrt{\eta_{h,\ell}^2 - 1} & \text{if } h^2\mu_{h,\ell}^2 \in (-6, 0). \end{cases}$$

We can further show that

$$\sqrt{\eta_\ell^2 - 1} = \frac{ih\mu_{h,\ell}\sqrt{1 - h^2\mu_{h,\ell}^2/12}}{1 + h^2\mu_{h,\ell}^2/6} \times \begin{cases} +1 & \text{if } h^2\mu_{h,\ell}^2 \in (-\infty, -6) \cup (0, 1/\varrho^2), \\ -1 & \text{if } h^2\mu_{h,\ell}^2 \in (-6, 0). \end{cases}$$

Combining it with (4.11) leads to (4.19). The same argument using (4.22) instead of (4.21) can show (4.20) and the proof is completed. □

4.3 Discrete PML operator and its matrix representation

Since the sesquilinear form $b_{\mathrm{PML}}(\cdot,\cdot)$ is coercive in $\widetilde{H}_0^1(\Omega_{\mathrm{PML}}) \times \widetilde{H}_0^1(\Omega_{\mathrm{PML}})$ as mentioned earlier in Subsection 2.2, for $g_h \in V_{\Gamma_0}^h$ the problem

$$b_{\mathrm{PML}}(u_h, v_h) = 0 \text{ for all } v_h \in \widetilde{V}_{\Omega_{\mathrm{PML}}}^h \tag{4.24}$$

with $u_h = g_h$ on Γ_0 has a unique solution $u_h \in V_{\Omega_{\mathrm{PML}}}^h$, where $\widetilde{V}_{\Omega_{\mathrm{PML}}}^h$ is the finite element space of functions v_h in $V_{\Omega_{\mathrm{PML}}}^h$ vanishing on Γ_0. Thus we can have an

extension operator $S^h : V^h_{\Gamma_0} \to V^h_{\Omega_{\rm PML}}$ defined by $S^h(g_h) = u_h$, the solution to the problem (4.24). Analogously to the continuous PML operator $T_{\rm PML}$ defined as the variational sense (2.6), we can define the discrete PML operator $T^h_{\rm PML} : V^h_{\Gamma_0} \to V^h_{\Gamma_0}$ such that

$$(T^h_{\rm PML} g_h, v_h)_{\Gamma_0} = -b_{\rm PML}(S^h(g_h), S^h(v_h)) \text{ for } v_h \in V^h_{\Gamma_0}. \tag{4.25}$$

Now, the finite element approximation $\tilde{u}_h \in V^h_{\tilde{\Omega}}$ to the continuous problem (2.4) is obtained by solving the problem

$$\tilde{b}(\tilde{u}_h, \tilde{v}_h) = (f, \tilde{v}_h)_{\Omega} \text{ for } \tilde{v}_h \in V^h_{\tilde{\Omega}}.$$

Then it can be shown that the restriction of $\tilde{u}_h$ to Ω is equal to the solution $u^{ex}_h \in V^h_{\Omega}$ to the problem

$$b^h(u^{ex}_h, v_h) = (f, v_h)_{\Omega} \text{ for } v_h \in V^h_{\Omega},$$

where

$$b^h(u_h, v_h) = (\nabla u_h, \nabla v_h)_{\Omega} - k^2 (u_h, v_h)_{\Omega} - (T^h_{\rm PML}(u_h), v_h)_{\Gamma_0}.$$

We remark that the discrete PML operator can be defined on Γ_j for $j = 1, \ldots, J-1$ for the absorbing boundary condition.

In the rest of this subsection we will study the matrix representation $\hat{T}^h_{\rm PML}$ of the discrete PML operator $T^h_{\rm PML}$. Noting that the triangulation $\mathcal{T}_{\Omega_{\rm PML}}$ on $\Omega_{\rm PML}$ is defined analogously to that for $\hat{\Omega}$ with N_p subintervals of $-\beta < x < 0$ ($N_p = \beta/h$), we define the subspaces $\mathcal{V}^h_{\Gamma,q}$ of finite element functions in $V^h_{\Omega_{\rm PML}}$ vanishing on all nodes outside of $\{-qh\} \times \Gamma$ for $q = 0, \ldots, N_p$ as done in the preceding section. Here we abuse the notation of the symbol $\mathcal{V}^h_{\Gamma,q}$ used for finite element spaces on $\hat{\Omega}$ since the meaning of the symbol is clear from context. Then the coordinate representation $\hat{u}_h = (\hat{u}_{h,0}, \ldots, \hat{u}_{h,N_p})$ of u_h to the problem (4.24) satisfies three term recurrence

$$-B_{\rm PML}\hat{u}_{h,q-1} + 2A_{\rm PML}\hat{u}_{h,q} - B_{\rm PML}\hat{u}_{h,q+1} = 0 \tag{4.26}$$

for $q = 1, 2, \ldots, N_p - 1$, and the boundary condition on Γ_0

$$\hat{u}_{h,0} = \hat{g}_h = \sum_{\ell=1}^{N_\Gamma} \hat{u}^{\ell}_{h,0} \hat{\phi}_{h,\ell} \tag{4.27}$$

where $\hat{g}_h$ is the coordinate representation of g_h and

$$A_{\rm PML} = \frac{1}{\sigma_0} A_x + \sigma_0 A_y - k^2 \sigma_0 A_0, \qquad B_{\rm PML} = \frac{1}{\sigma_0} B_x + \sigma_0 B_y - k^2 \sigma_0 B_0.$$

Since $1 + h^2 \sigma_0^2 \mu^2_{h,\ell}/6 \neq 0$ for all ℓ, we can prove that the generalized eigenvalue problem for the matrices $A_{\rm PML}$ and $B_{\rm PML}$ has a full set of eigenvectors in the same way as in Lemma 4.2.

Lemma 4.5 *The generalized eigenvalue problem*

$$A_{\mathrm{PML}}\hat{\phi} = \eta B_{\mathrm{PML}}\hat{\phi} \tag{4.28}$$

has the eigenvectors $\hat{\phi}_{h,\ell}$ *for eigenvalues*

$$\eta_{p,\ell} = \left(1 - \frac{h^2\sigma_0^2\mu_{h,\ell}^2}{3}\right)\left(1 + \frac{h^2\sigma_0^2\mu_{h,\ell}^2}{6}\right)^{-1}$$

for $\ell = 1, 2, \ldots, N_\Gamma$.

The characteristic equation $\xi^2 - 2\eta_{p,\ell}\xi + 1 = 0$ has two distinct solutions, one of which is smaller than 1 in magnitude, which is the same effect as PML does in the continuous level, that is, transforming propagating modes to evanescent modes. More precisely, we set

$$\xi_{p,\ell} = \eta_{p,\ell} + \mathrm{sgn}(\mu_{h,\ell}^2)\sqrt{\eta_{p,\ell}^2 - 1}^{\,*}. \tag{4.29}$$

Here $\sqrt{z}^{\,*}$ is the square root function with branch cut $0 \le \arg(z) < 2\pi$ satisfying

$$\sqrt{z}^{\,*} = \sqrt{z} \text{ for } \Im(z) \ge 0 \quad \text{and} \quad \sqrt{z}^{\,*} = -\sqrt{z} \text{ for } \Im(z) < 0. \tag{4.30}$$

By observing that

$$\Im(\eta_{p,\ell}) > 0 \text{ for } \mu_{h,\ell}^2 < 0 \quad \text{and} \quad \Im(\eta_{p,\ell}) < 0 \text{ for } \mu_{h,\ell}^2 > 0,$$

$$|z - \sqrt{z^2-1}^{\,*}| < 1 \text{ for } \Im(z) > 0 \quad \text{and} \quad |z + \sqrt{z^2-1}^{\,*}| < 1 \text{ for } \Im(z) < 0,$$

we find that $\xi_{p,\ell}$ is the solution to the characteristic equation such that $|\xi_{p,\ell}| < 1$. Moreover, $\xi_{p,\ell}$ can be written as

$$\xi_{p,\ell} = \frac{1 - \frac{h^2\mu_{h,\ell}^2\sigma_0^2}{3} + ih\mu_{h,\ell}\sigma_0\sqrt{1 - \frac{h^2\mu_{h,\ell}^2\sigma_0^2}{12}}}{1 + \frac{h^2\mu_{h,\ell}^2\sigma_0^2}{6}}. \tag{4.31}$$

Now we have the solution formula satisfying three term recurrence (4.26), which is given in the following lemma.

Lemma 4.6 *Suppose that* $-B_{\mathrm{PML}}\hat{u}_{h,q-1} + 2A_{\mathrm{PML}}\hat{u}_{h,q} - B_{\mathrm{PML}}\hat{u}_{h,q+1} = 0$ *for* $q = 1, 2, \ldots, N_p - 1$. *Then* $\hat{u}_{h,q}$ *for* $q = 0, 1, \ldots, N_p$ *is of the form*

$$\hat{u}_{h,q} = \sum_{\ell=1}^{N_\Gamma}(\alpha_\ell\xi_{p,\ell}^q + \beta_\ell\xi_{p,\ell}^{-q})\hat{\phi}_{h,\ell} \tag{4.32}$$

for constants $\alpha_\ell, \beta_\ell \in \mathbb{C}$.

Now, we are ready to find the matrix representation $\hat{T}_{\mathrm{PML}}^h$ of the discrete DtN operator T_{PML}^h.

Lemma 4.7 *Let T^h_{PML} be the discrete* PML *operator defined by* (4.25)*. Then for $\hat{u}_{h,0}$ expanded as* (4.27) *the matrix representation of T^h_{PML} is given by*

$$\hat{T}^h_{\mathrm{PML}}\hat{u}_{h,0} = \sum_{\ell=1}^{N_\Gamma} \Lambda_{p,\ell}\hat{u}^\ell_{h,0}\hat{\phi}_{h,\ell}, \tag{4.33}$$

where

$$\Lambda_{p,\ell} = i\mu_{h,\ell}\sqrt{1-\frac{h^2\sigma_0^2\mu_{h,\ell}^2}{12}}\,\frac{1-\xi_{p,\ell}^{2N_p}}{1+\xi_{p,\ell}^{2N_p}}.$$

Proof We first note that the coordinate representation $\hat{u}_h$ of the solution $u_h \in V^h_{\Omega_{\mathrm{PML}}}$ to the problem (4.24) satisfies (4.26) and hence the homogeneous Neumann boundary condition on $\Gamma_\beta = \{-\beta\}\times\Gamma$, $-B_{\mathrm{PML}}\hat{u}_{h,N_p-1} + A_{\mathrm{PML}}\hat{u}_{h,N_p} = 0$, can be read as $\alpha_\ell\xi_{p,\ell}^{N_p} - \beta_\ell\xi_{p,\ell}^{-N_p} = 0$ for $\ell = 1,\dots,N_\Gamma$ by Lemma 4.6. Thus we can use the boundary condition (4.27) to obtain

$$\alpha_\ell = \frac{\hat{u}^\ell_{h,0}}{1+\xi_{p,\ell}^{2N_p}} \quad\text{and}\quad \beta_\ell = \frac{\xi_{p,\ell}^{2N_p}\hat{u}^\ell_{h,0}}{1+\xi_{p,\ell}^{2N_p}}.$$

Since the matrix representation $\hat{T}^h_{\mathrm{PML}}$ of the discrete DtN operator T^h_{PML} satisfies

$$M_\Gamma\hat{T}^h_{\mathrm{PML}}\hat{u}_{h,0} = -A_{\mathrm{PML}}\hat{u}_{h,0} + B_{\mathrm{PML}}\hat{u}_{h,1},$$

by using Lemma 4.6 we have

$$\begin{aligned} M_\Gamma\hat{T}^h_{\mathrm{PML}}\hat{u}_{h,0} &= \sum_{\ell=1}^{N_\Gamma}\frac{1}{2}(\xi_{p,\ell}-\xi_{p,\ell}^{-1})(\alpha_\ell-\beta_\ell)B_{\mathrm{PML}}\hat{\phi}_{h,\ell} \\ &= \sum_{\ell=1}^{N_\Gamma}\frac{1}{2}(\xi_{p,\ell}-\xi_{p,\ell}^{-1})\frac{1-\xi_{p,\ell}^{2N_p}}{1+\xi_{p,\ell}^{2N_p}}\hat{u}^\ell_{h,0}B_{\mathrm{PML}}\hat{\phi}_{h,\ell}. \end{aligned} \tag{4.34}$$

A simple computation using (4.31) shows that

$$\frac{1}{2}(\xi_{p,\ell}-\xi_{p,\ell}^{-1})B_{\mathrm{PML}}\hat{\phi}_{h,\ell} = i\mu_{h,\ell}\sqrt{1-\frac{h^2\sigma_0^2\mu_{h,\ell}^2}{12}}\,M_\Gamma\hat{\phi}_{h,\ell}$$

and hence we have the formula (4.33) for the discrete PML operator. □

4.4 Error propagation in the discrete level

We consider the discrete local problem (4.3) posed in the reference domain $\hat{\Omega}$ to find $\hat{u}_h = (\hat{u}_{h,0}\dots,\hat{u}_{h,N_s}) \in \mathbb{C}^{N_\Gamma\times(N_s+1)}$ satisfying

$$-B\hat{u}_{h,q-1} + 2A\hat{u}_{h,q} - B\hat{u}_{h,q+1} = 0 \text{ for } q = 1,2,\dots,N_s-1 \tag{4.35}$$

and the boundary conditions

$$\frac{\partial^h\hat{u}_h}{\partial\nu} = \hat{T}^h_{\mathrm{PML}}\hat{u}_{h,0} + \hat{\gamma}^{\mathrm{L}}_{\mathrm{in}} \text{ on } \Gamma_{\mathrm{L}} \quad\text{and}\quad \frac{\partial^h\hat{u}_h}{\partial\nu} = \hat{T}^h_{\mathrm{PML}}\hat{u}_{h,N_s} + \hat{\gamma}^{\mathrm{R}}_{\mathrm{in}} \text{ on } \Gamma_{\mathrm{R}} \tag{4.36}$$

with $\hat{\gamma}_{\text{in}}^{\text{L}}$ and $\hat{\gamma}_{\text{in}}^{\text{R}} \in \mathbb{C}^{N_\Gamma}$. Once the problem (4.35)-(4.36) is solved, the outgoing data corresponding to (3.10) are defined by

$$\hat{\gamma}_{\text{out}}^{\text{L}} = -\frac{\partial^h \hat{u}_h}{\partial \nu} - \hat{T}_{\text{PML}}^h \hat{u}_{h,0} \quad \text{and} \quad \hat{\gamma}_{\text{out}}^{\text{R}} = -\frac{\partial^h \hat{u}_h}{\partial \nu} - \hat{T}_{\text{PML}}^h \hat{u}_{h,N_s}. \tag{4.37}$$

Remark 4.8 *We will assume $J_B = \varnothing$ in the analysis of the rest of the paper since we rarely encounter the case $J_B \neq \varnothing$, that is $h^2\mu_{h,\ell}^2 = -6$, in the most common situation in reality. However, one can easily show the double sweep DDM for the modes with $\ell \in J_B \neq \varnothing$ converges as well.*

Recalling the solution formula (4.16) under the assumption $J_B = \varnothing$, for $\gamma_{\text{in}}^{\text{L}} = \sum_{\ell=1}^{N_\Gamma} \gamma_{\text{in},\ell}^{\text{L}} \hat{\phi}_{h,\ell}$ and $\gamma_{\text{in}}^{\text{R}} = \sum_{\ell=1}^{N_\Gamma} \gamma_{\text{in},\ell}^{\text{R}} \hat{\phi}_{h,\ell}$, and using the matrix representations of discrete normal derivatives and discrete PML operator given in Lemma 4.4 and Lemma 4.7 we solve the equations (4.36) for α_ℓ and β_ℓ, and then use them in the equations (4.37) for the outgoing data as done for (3.12) in the continuous level problem to obtain that

$$\begin{bmatrix} \hat{\gamma}_{\text{out},\ell}^{\text{L}} \\ \hat{\gamma}_{\text{out},\ell}^{\text{R}} \end{bmatrix} = \begin{bmatrix} \varepsilon_{h,\ell} & \zeta_{h,\ell} \\ \zeta_{h,\ell} & \varepsilon_{h,\ell} \end{bmatrix} \begin{bmatrix} \hat{\gamma}_{\text{in},\ell}^{\text{L}} \\ \hat{\gamma}_{\text{in},\ell}^{\text{R}} \end{bmatrix}, \tag{4.38}$$

where

$$\zeta_{h,\ell} = \frac{(1 - Q_{h,\ell}^2)\xi_{h,\ell}^{N_s}}{1 - Q_{h,\ell}^2 \xi_{h,\ell}^{2N_s}}, \qquad \varepsilon_{h,\ell} = \frac{(1 - \xi_{h,\ell}^{2N_s})Q_{h,\ell}}{1 - Q_{h,\ell}^2 \xi_{h,\ell}^{2N_s}}. \tag{4.39}$$

Here $Q_{h,\ell}$ is the discrete reflection coefficients for the ℓ-th mode defined by

$$Q_{h,\ell} = \frac{-\Lambda_{h,\ell} + \Lambda_{p,\ell}}{-\Lambda_{h,\ell} - \Lambda_{p,\ell}} = \frac{\sqrt{1 - \frac{h^2\mu_{h,\ell}^2}{12}} - \sqrt{1 - \frac{h^2\sigma_0^2\mu_{h,\ell}^2}{12}} \frac{1-\xi_{p,\ell}^{2N_p}}{1+\xi_{p,\ell}^{2N_p}}}{-\sqrt{1 - \frac{h^2\mu_{h,\ell}^2}{12}} - \sqrt{1 - \frac{h^2\sigma_0^2\mu_{h,\ell}^2}{12}} \frac{1-\xi_{p,\ell}^{2N_p}}{1+\xi_{p,\ell}^{2N_p}}}. \tag{4.40}$$

Denoting $z = h\mu_{h,\ell}$, by Taylor's theorem one can show that the discrete reflection coefficient $Q_{h,\ell}$ satisfies the asymptotic behavior

$$Q_{h,\ell} = \frac{1}{48}(1 - \sigma_0^2)z^2 - \xi_{p,\ell}^{2N_p} + O(\|(z^2, \xi_{p,h}^{2N_p})\|^2) \tag{4.41}$$

for small z and $\xi_{p,h}^{2N_p}$.

From now on we will estimate the entries $\varepsilon_{h,\ell}$ and $\zeta_{h,\ell}$ of the error propagation matrix as well as the discrete reflection coefficient $Q_{h,\ell}$ in more detail. Since $Q_{h,\ell}$ depends on $|\xi_{p,\ell}|^{2N_p}$, we begin by estimating $|\xi_{p,\ell}|^{2N_p}$. We let $w = h\mu_{h,\ell}\sigma_0$ and write $\xi_{p,\ell}$ in (4.31) as the function of w

$$\xi_{p,\ell} = \chi(w) = \frac{1 - \frac{w^2}{3} + iw\sqrt{1 - \frac{w^2}{12}}}{1 + \frac{w^2}{6}}.$$

Clearly, we have $|\chi(w)| < 1$ from the definition of $\xi_{p,\ell} = \chi(w)$ and the asymptotic behavior of $\chi(w)$ holds

$$\chi(w) = 1 + iw + \mathcal{R}_\chi(w) \tag{4.42}$$

with $\mathcal{R}_\chi(w) = O(w^2)$ for small w.

Due to the convergence theory of finite element eigenvalues, for small $0 < \epsilon_* < \mu_{\min}$ we can find a positive constant h_0 such that if $0 < h \le h_0$, then the compact sets $[0, \mu_N - \epsilon_*]$ and $i[0, \tilde{\mu}_{N+1} - \epsilon_*]$ do not have discrete wavenumbers $\mu_{h,\ell}$ of any modes. The decaying property of $\xi_{p,\ell}^{2N_p}$ is presented in the following lemma, whose proof will be given in Appendix.

Lemma 4.9 *For any $\epsilon > 0$ there exists $0 < h_1 \le h_0$ such that if $0 < h \le h_1$, then*

$$|\xi_{p,\ell}|^{2N_p} \le (1+\epsilon)e^{-2\sigma_\mu \beta}$$

for all $\ell = 1, \ldots, N_\Gamma$.

From Lemma 4.9 and the asymptotic behavior (4.41) of $Q_{h,\ell}$ for small z and $\xi_{p,\ell}^{2N_p}$, we can easily show that the discrete reflection coefficient $Q_{h,\ell}$ can be asymptotically reduced to a constant multiple of $e^{-2\sigma_\mu \beta}$ provided that h and z are small enough.

Lemma 4.10 *Let $\delta_z := \sqrt{\left|48/(1-\sigma_0^2)\right|}e^{-\sigma_\mu \beta}$. Then it holds that*

$$|Q_{h,\ell}| \lesssim e^{-2\sigma_\mu \beta} \tag{4.43}$$

for $0 < h \le h_1$ and $|h\mu_{h,\ell}| \le \delta_z$.

We may need to choose small h so that the discrete reflection coefficients $Q_{h,\ell}$ for all discrete propagating modes fulfill the condition (4.43), that is, $h\mu_{h,\ell} \le \delta_z$ for all $\mu_{h,\ell} > 0$, however we do not introduce another bound of h for simple presentation and assume that if $h \le h_1$, then $h\mu_{h,\ell} \le \delta_z$ for $\mu_{h,\ell} > 0$. We define $\ell_* > N$ such that if $\ell \le \ell_*$, then $|h\mu_{h,\ell}| \le \delta_z$ but $|h\mu_{h,\ell}| > \delta_z$ otherwise, and note that ℓ_* increases to N_Γ as h tends toward zero.

Remark 4.11 *We further assume that h_1 is small enough so that the ℓ-th evanescent modes corresponding to $\mu_{h,\ell}$ for $\ell > \ell_*$ are sufficiently small, $|\xi_{h,\ell}^{N_s}| \ll 1$. Indeed, as inferred from the graph of $\xi = \eta \pm \sqrt{\eta^2 - 1}$ as a function of $x = (h\mu)^2 < 0$ for evanescent modes in Fig. 2, we have $|\xi_{h,\ell}| < C_\xi < 1$ for some C_ξ depending only on δ_z when $|h\mu_{h,\ell}| \ge \delta_z$, from which it follows that $|\xi_{h,\ell}^{N_s}| = |\xi_{h,\ell}^{H/h}| \to 0$ as $h \to 0$. Thus we can assume that h_1 is small enough so that if $0 < h \le h_1$, then $|\xi_{h,\ell}^{N_s}| \ll 1$ for the ℓ-th evanescent modes with $\ell > \ell_*$. The exponential decay of $\xi_{h,\ell}^{N_s}$ is engaged significantly in the error analysis because the discrete reflection coefficients $Q_{h,\ell}$, satisfying the asymptotic formula*

$$|Q_{h,\ell}| \approx \left| \frac{\sqrt{1+x} - \sqrt{1+x\sigma_0^2}}{\sqrt{1+x} + \sqrt{1+x\sigma_0^2}} \right| \tag{4.44}$$

with $x = (h\mu_{h,\ell})^2/12$ resulting from (4.40) for sufficiently small $\xi_{p,\ell}^{2N_p}$, get worse with increasing ℓ.

Next, we estimate the entries of the error propagation matrix for small $z = h\mu_{h,\ell}$.

Lemma 4.12 *Assume that* $0 < h \le h_1$ *and* $\ell \le \ell_*$. *Then it holds that*

$$|\varepsilon_{h,\ell}| \le C_d e^{-2\sigma_\mu \beta}$$
$$|\zeta_{h,\ell}| \le 1 + C_d e^{-4\sigma_\mu \beta}$$

for a positive generic constant C_d.

Proof Due to the fact that $|Q_{h,\ell}| \le Ce^{-2\sigma_\mu\beta} \ll 1$ and $|\xi_{h,\ell}^{N_s}| \le 1$, we use the Taylor expansion with respect to $Q_{h,\ell}$ to obtain

$$\left|\frac{1-\xi_{h,\ell}^{2N_s}}{1-Q_{h,\ell}^2\xi_{h,\ell}^{2N_s}}\right| = \left|(1-\xi_{h,\ell}^{2N_s}) + O(Q_{h,\ell}^2)\right| \le C,$$
$$\left|\frac{1-Q_{h,\ell}^2}{1-Q_{h,\ell}^2\xi_{h,\ell}^{2N_s}}\right| = \left|1 + O(Q_{h,\ell}^2)\right| \le 1 + Ce^{-4\sigma_\mu\beta}$$

for a generic constant C. Then, the required estimates result from the definition (4.39) of $\varepsilon_{h,\ell}$ and $\zeta_{h,\ell}$ and the above inequalities. □

The next lemma deals with the case of $|z| > \delta_z$, which can be proved in the similar way by using $|\xi_{h,\ell}^{N_s}| \ll 1$ instead of $|Q_{h,\ell}| \ll 1$.

Lemma 4.13 *Assume that* $0 < h \le h_1$ *and* $\ell > \ell_*$. *Then it holds that*

$$|\varepsilon_{h,\ell}| \le (1 + C|\xi_{h,\ell}^{2N_s}|)|Q_{h,\ell}|,$$
$$|\zeta_{h,\ell}| \le C|\xi_{h,\ell}^{2N_s}|$$

for a generic constant C.

5 Double sweep for the finite element problem

In this section we analyze the convergence of the double sweep DDM applied to the Helmholtz equation discretized by the finite element method.

5.1 Embedding operators

Let I_j^{in} and I_{Γ_j} be the index sets of nodes on $\bar{\Omega}_j \setminus (\overline{\Gamma}_j \cup \overline{\Gamma}_{j-1})$ and $\bar{\Gamma}_j$, respectively. Then every function in $V_{\Omega_j}^h$ is obtained by restricting a function in V_Ω^h to Ω_j, that is, for $v_h \in V_{\Omega_j}^h$, we can write v_h as a linear combination of nodal basis functions φ_i in V_Ω^h,

$$v_h = \sum_{i \in I_j^{in}} v_{h,i}\varphi_i|_{\Omega_j} + \sum_{i \in I_{\Gamma_j}} v_{h,i}\varphi_i|_{\Omega_j} + \sum_{i \in I_{\Gamma_{j-1}}} v_{h,i}\varphi_i|_{\Omega_j} \tag{5.1}$$

for $v_{h,i} \in \mathbb{C}$.

We remark that the double sweep DDM in the continuous level studied in Section 3 produces a convergent sequence of approximate solutions, that lie in $H^1(\Omega_j)$ locally but are discontinuous on the interfaces Γ_j in general. Because local

solutions of each iterate are defined in mutually disjoint subdomains, they can hold data γ_j^{L} and γ_j^{R} to be transferred to neighboring subdomains on interfaces. In the double sweep DDM for the discretized problem, we will find a sequence u_h^m in $V_\Omega^h \subset H^1(\Omega)$ converging to u_h^{ex}, that has one trace instead of two different traces on Γ_j, but holds correct outgoing data for the next iterate. We can do this by updating each iterate in a special way. To this end, we need restricted embeddings of $V_{\Omega_j}^h$ into V_Ω^h. We define $\mathbb{E}_j^F : V_{\Omega_j}^h \to V_\Omega^h$ for $1 \le j \le J-1$ and $\mathbb{E}_j^B : V_{\Omega_j}^h \to V_\Omega^h$ for $1 \le j \le J$ by

$$\mathbb{E}_j^F(v_h) = \begin{cases} \displaystyle\sum_{i \in I_j^{in}} v_{h,i}\varphi_i + \sum_{i \in I_{\Gamma_j}} v_{h,i}\varphi_i + \sum_{i \in I_{\Gamma_{j-1}}} v_{h,i}\varphi_i & \text{for } j = 1, \\ \displaystyle\sum_{i \in I_j^{in}} v_{h,i}\varphi_i + \sum_{i \in I_{\Gamma_j}} v_{h,i}\varphi_i & \text{for } 2 \le j \le J-1, \end{cases}$$

$$\mathbb{E}_j^B(v_h) = \sum_{i \in I_j^{in}} v_{h,i}\varphi_i + \sum_{i \in I_{\Gamma_{j-1}}} v_{h,i}\varphi_i \qquad \text{for } 1 \le j \le J$$

for $v_h \in V_{\Omega_j}^h$ of (5.1). We note that the operator $\mathbb{E}_j^F$ defined in $V_{\Omega_j}^h$ keeps the boundary values on Γ_j but not Γ_{j-1} (except for $j = 1$), which will be used for the forward sweep. In contrast, the operator $\mathbb{E}_j^B$ defined in $V_{\Omega_j}^h$ keeps the boundary values on Γ_{j-1} but not Γ_j, which will be employed for the backward sweep.

5.2 Algorithm for the convergence theory

In this subsection we discuss the double sweep DDM algorithm for the discrete problem. We first note that the discrete incoming data $\gamma_{h,j}^{\mathrm{L}}$ and $\gamma_{h,j}^{\mathrm{R}}$ corresponding to γ_j^{L} and γ_j^{R} for the continuous level can be computed theoretically as follows. Assuming that $u_h \in V_\Omega^h$ is known and denoting by $u_{h,j}$ the restriction of u_h to Ω_j (the symbol $u_{h,q}$ has been used for a function in $\mathcal{V}_{\Gamma,q}^h$ in Section 4 but $u_{h,j}$ in this section represents a function in $V_{\Omega_j}^h$) so that $u_{h,j-1} \in V_{\Omega_{j-1}}^h$ and $u_{h,j+1} \in V_{\Omega_{j+1}}^h$ are given, the incoming data coming into Ω_j are determined by

$$\gamma_{h,j}^{\mathrm{L}} = \frac{\partial^h u_{h,j-1}}{\partial \nu_j} - T_{\mathrm{PML}}^h(u_{h,j-1}), \quad \gamma_{h,j}^{\mathrm{R}} = \frac{\partial^h u_{h,j+1}}{\partial \nu_j} - T_{\mathrm{PML}}^h(u_{h,j+1}) \tag{5.2}$$

on Γ_{j-1} and Γ_j, respectively.

In the double sweep iteration, we denote by $\boldsymbol{\gamma}_h^{\mathrm{R},m}$ and $\boldsymbol{\gamma}_h^{\mathrm{L},m}$ the discrete boundary data corresponding to $\boldsymbol{\gamma}^{\mathrm{R},m}$ and $\boldsymbol{\gamma}^{\mathrm{L},m}$, respectively,

$$\boldsymbol{\gamma}_h^{\mathrm{R},m} = (\gamma_{h,1}^{\mathrm{R},m}, \gamma_{h,2}^{\mathrm{R},m}, \ldots, \gamma_{h,J-1}^{\mathrm{R},m}) \in G_h^{\mathrm{R}}, \quad \boldsymbol{\gamma}_h^{\mathrm{L},m} = (\gamma_{h,2}^{\mathrm{L},m}, \gamma_{h,3}^{\mathrm{L},m}, \ldots, \gamma_{h,J}^{\mathrm{L},m}) \in G_h^{\mathrm{L}},$$

where $G_h^{\mathrm{R}} := \prod_{j=1}^{J-1} V_{\Gamma_j}^h$ and $G_h^{\mathrm{L}} := \prod_{j=2}^{J} V_{\Gamma_j}^h$. Then the discrete double sweep DDM keeping correct traces on the interfaces by using proper embedding operators defined in Subsection 5.1 can be given as Algorithm 2.

In this algorithm, it is not computationally cheap to extract incoming data $\gamma_{h,j}^{\mathrm{L},m}$ and $\gamma_{h,j}^{\mathrm{R},m}$ by using the formulas (5.2) since the discrete PML operators are involved. However, Algorithm 2 is needed as it plays a crucial role in the convergence analysis. An efficient way to avoid computing $\gamma_{h,j}^{\mathrm{L},m}$ and $\gamma_{h,j}^{\mathrm{R},m}$ directly in numerical implementations is discussed in [19].

Algorithm 2 Double sweep DDM of the discrete level problem

Set $m = 0$ and choose any $u_h^0 \in V_\Omega^h$.
Compute $\boldsymbol{\gamma}_h^{\mathrm{R},0} = (\gamma_{h,1}^{\mathrm{R},0}, \ldots, \gamma_{h,J-1}^{\mathrm{R},0})$ from u_h^0.
while the residual is larger than given tolerance **do**
  Set $u_h^{m+1} \leftarrow u_h^m$. ▷ Forward sweep
  for $j = 1, 2, \ldots, J-1$ **do**
    i. Compute $\gamma_{h,j}^{\mathrm{L},m}$ from $u_{h,j-1}^{m+1}$ with $\gamma_{h,1}^{\mathrm{L},m} = 0$.
    ii. Solve the local problem for $w_{h,j} \in V_{\Omega_j}^h$

$$b_j^h(w_{h,j}, v_h) = (f, v_h)_{\Omega_j} + (\gamma_{h,j}^{\mathrm{L},m}, v_h)_{\Gamma_{j-1}} + (\gamma_{h,j}^{\mathrm{R},m}, v_h)_{\Gamma_j} \text{ for } v_h \in V_{\Omega_j}^h. \tag{5.3}$$

    iii. Update the values of u_h^{m+1} corresponding to nodes in $I_j^{in} \cup I_{\Gamma_j}$ by using $\mathbb{E}_j^F(w_{h,j})$.
  end for
  Compute $\gamma_{h,J}^{\mathrm{L},m}$ from $u_{h,J-1}^{m+1}$. ▷ backward sweep
  for $j = J, J-1, \ldots, 1$ **do**
    i. Compute $\gamma_{h,j}^{\mathrm{R},m+1}$ from $u_{h,j+1}^{m+1}$ with $\gamma_{h,J}^{\mathrm{R},m+1}$ being ignored.
    ii. Solve the local problem (5.3) for $w_{h,j} \in V_{\Omega_j}^h$ with $\gamma_{h,j}^{\mathrm{R},m}$ replaced by $\gamma_{h,j}^{\mathrm{R},m+1}$.
    iii. Update the values of u_h^{m+1} corresponding to nodes $I_j^{in} \cup I_{\Gamma_{j-1}}$ by using $\mathbb{E}_j^B(w_{h,j})$.
  end for
  Set $m \leftarrow m + 1$.
end while

5.3 Convergence of the discrete double sweep DDM

In this subsection, we analyze the convergence of the discrete double sweep DDM as the main result based on the reflection coefficients studied in Subsection 4.4. The main result is that the double sweep process can be viewed as a contraction mapping of the boundary data coming from the right boundaries of subdomains and its contraction factor is determined by the maximal reflection coefficient depending on which reflection coefficient of $\ell \le \ell_*$ or $\ell > \ell_*$ is dominant. It turns out that the number of iterations can also increase logarithmically with the number of subdomains if the reflection coefficient of $\ell \le \ell_*$ including propagating modes is larger than the other.

As the continuous level problem, we estimate the discrete forward sweep operator $\mathbb{F}_h : \boldsymbol{\gamma}_h^{\mathrm{R},m} \mapsto \boldsymbol{\gamma}_h^{\mathrm{L},m}$ and the analogous one for backward sweep $\mathbb{B}_h : \boldsymbol{\gamma}_h^{\mathrm{L},m} \mapsto \boldsymbol{\gamma}_h^{\mathrm{R},m+1}$. We proceed the convergence analysis for the double sweep operator by splitting two cases depending on whether the mode index ℓ is larger than ℓ_*. For $\gamma_h \in V_\Gamma^h$ we let $\gamma_{h,\le}$ and $\gamma_{h,>}$ be consisting of the Fourier components of γ_h with $\ell \le \ell_*$ and $\ell > \ell_*$, respectively. Also, $\boldsymbol{\gamma}_{h,\le}$ and $\boldsymbol{\gamma}_{h,>}$ are analogously defined for vector functions.

5.3.1 The case of $\ell \le \ell_*$

In this case, we use Lemma 4.12 instead of Lemma 3.2 in the arguments used for verifying that $\mathbb{B} \circ \mathbb{F}$ is a contraction mapping in Subsection 3.3, and hence we are led to the following lemma showing that the double sweep operator in the finite element problem serves as a contraction mapping for modes of $\ell \le \ell_*$.

Lemma 5.1 *Assume that $h < h_1$. Let c_{0h} be a positive constant such that $c_{0h}(J-1) < 1$ as well as $(2C_d c_{0h}^2 + C_d^2 c_{0h}^4)(J-1) < 1$. If the* PML *parameters are chosen such that $e^{-2\sigma_\mu \beta} < c_{0h}$, then for modes such that $|h\mu_{h,\ell}| < \delta_z$ it holds that*

$$\|\boldsymbol{\gamma}_{h,\leq}^{\mathrm{L},m}\|_{\dot{H}_h^{-1/2}(\Gamma)} \leq C_\delta (J-1) e^{-2\sigma_\mu \beta} \|\boldsymbol{\gamma}_{h,\leq}^{\mathrm{R},m}\|_{\dot{H}_h^{-1/2}(\Gamma)},$$
$$\|\boldsymbol{\gamma}_{h,\leq}^{\mathrm{R},m+1}\|_{\dot{H}_h^{-1/2}(\Gamma)} \leq C_\delta (J-1) e^{-2\sigma_\mu \beta} \|\boldsymbol{\gamma}_{h,\leq}^{\mathrm{R},m}\|_{\dot{H}_h^{-1/2}(\Gamma)}$$

for some generic constant $C_\delta > 0$ that may depend only on k.

Proof The estimates are the results for the discrete level problem analogous to those for the continuous level problem given in Lemma 3.4, Lemma 3.5 and Lemma 3.6. In order to establish the estimates, we have only to follow the same lines used for the three lemmas for the continuous level problem by using Lemma 4.12 instead of Lemma 3.2 once the discrete version of the stability corresponding to (3.23) is provided. Thus, it suffices to prove

$$\|\gamma_{h,J-1,\leq}^{\mathrm{R},m+1}\|_{\dot{H}_h^{-1/2}(\Gamma)} \lesssim \|\gamma_{h,J,\leq}^{\mathrm{L},m}\|_{\dot{H}_h^{-1/2}(\Gamma)}.$$

To this end, let $\Omega_{\mathrm{PML}}^J = (x_{J-1} - \beta, x_{J-1}) \times \Gamma$ be the PML region for the local problem posed on Ω_J and $\tilde{u}_{h,J}$ be the solution to the problem in $\widetilde{\Omega}_J$ defined by Ω_J attached to Ω_{PML}^J with $\gamma_{h,J,\leq}^{\mathrm{L},m}$ given on the transmission condition on Γ_{J-1},

$$a_J(\tilde{u}_{h,J}, \phi_h) + b_{\mathrm{PML}}^J(\tilde{u}_{h,J}, \phi_h) = (\gamma_{h,J,\leq}^{\mathrm{L},m}, \phi_h)_{\Gamma_{J-1}} \text{ for all } \phi_h \in V_{\widetilde{\Omega}_J}^h, \tag{5.4}$$

where $b_{\mathrm{PML}}^J(\cdot,\cdot)$ is defined analogously to $b_{\mathrm{PML}}(\cdot,\cdot)$ and $V_{\widetilde{\Omega}_J}^h$ is the finite element space on $\widetilde{\Omega}_J$. Then the restriction $u_{h,J}$ of $\tilde{u}_{h,J}$ to Ω_J satisfies the stability result

$$\|u_{h,J}\|_{H^1(\Omega_J)} \lesssim \|\gamma_{h,J,\leq}^{\mathrm{L},m}\|_{\dot{H}_h^{-1/2}(\Gamma)}. \tag{5.5}$$

Using (4.18), (4.25), (5.4) and (5.5), we have

$$\begin{aligned}
&|(\gamma_{h,J-1,\leq}^{\mathrm{R},m+1}, \phi_h)_{\Gamma_{J-1}}| = |a_J(\tilde{u}_{h,J}, \phi_h) - b_{\mathrm{PML}}^J(\tilde{u}_{h,J}, \phi_h)| \\
&= |2a_J(u_{h,J}, \phi_h) - (\gamma_{h,J,\leq}^{\mathrm{L},m}, \phi_h)_{\Gamma_{J-1}}| \lesssim \|\gamma_{h,J,\leq}^{\mathrm{L},m}\|_{\dot{H}_h^{-1/2}(\Gamma)} \|\phi_h\|_{H^1(\Omega)}.
\end{aligned}$$

Utilizing a bounded extension operator from $V_{\Gamma_{J-1}}^h$ to $V_{\widetilde{\Omega}_J}^h$ and (4.2), we obtain that

$$\|\gamma_{h,J-1,\leq}^{\mathrm{R},m+1}\|_{\dot{H}_h^{-1/2}(\Gamma)} \lesssim \sup_{0 \neq \phi_h \in V_{\Gamma_{J-1}}^h} \frac{|(\gamma_{h,J-1,\leq}^{\mathrm{R},m+1}, \phi_h)_{\Gamma_{J-1}}|}{\|\phi_h\|_{H^{1/2}(\Gamma)}} \lesssim \|\gamma_{h,J,\leq}^{\mathrm{L},m}\|_{\dot{H}_h^{-1/2}(\Gamma)},$$

which completes the proof. □

5.3.2 The case of $\ell > \ell_*$

In this case we have $|\zeta_{h,\ell}| \ll 1$ by Remark 4.11 and Lemma 4.13. We write the forward sweep operator as a matrix form

$$\boldsymbol{\gamma}_{h,\ell}^{\mathrm{L},m} = \varepsilon_{h,\ell}\Xi_{h,\ell}\boldsymbol{\gamma}_{h,\ell}^{\mathrm{R},m}, \tag{5.6}$$

where $\Xi_{h,\ell}$ is the $(J-1)\times(J-1)$ Toeplitz matrix given by (3.21) with ζ_n replaced by $\zeta_{h,\ell}$.

For the analysis of the backward sweep operator, we need to investigate error propagation arising in Ω_J. We conduct the similar computation used for (4.38) with the transmission condition on Γ_J replaced by the homogeneous Neumann boundary condition to obtain

$$\gamma_{h,J-1,\ell}^{\mathrm{R},m+1} = \frac{Q_{h,\ell} + \xi_{h,\ell}^{2N_s}}{1 + Q_{h,\ell}\xi_{h,\ell}^{2N_s}}\gamma_{h,J,\ell}^{\mathrm{L},m} := \tilde{\varepsilon}_{h,\ell}\gamma_{h,J,\ell}^{\mathrm{L},m}$$

with $|\tilde{\varepsilon}_{h,\ell}| = |Q_{h,\ell}| + O(|\xi_{h,\ell}^{2N_s}|)$. Inductively, it can be shown that

$$\boldsymbol{\gamma}_{h,\ell}^{\mathrm{R},m+1} = diag(\varepsilon_{h,\ell},\ldots,\varepsilon_{h,\ell},\tilde{\varepsilon}_{h,\ell})\Xi_{h,\ell}^{\top}\boldsymbol{\gamma}_{h,\ell}^{\mathrm{L},m}. \tag{5.7}$$

Combining (5.6) and (5.7), we get the action of the double sweep operator

$$\boldsymbol{\gamma}_{h,\ell}^{\mathrm{R},m+1} = diag(\varepsilon_{h,\ell},\ldots,\varepsilon_{h,\ell},\tilde{\varepsilon}_{h,\ell})\varepsilon_{h,\ell}\Xi_{h,\ell}^{\top}\Xi_{h,\ell}\boldsymbol{\gamma}_{h,\ell}^{\mathrm{R},m}.$$

Let us define

$$q_Q = \max_{\ell>\ell_*}\{|\varepsilon_{h,\ell}|\|\Xi_{h,\ell}\|_{\ell^2},\quad |\tilde{\varepsilon}_{h,\ell}|\|\Xi_{h,\ell}\|_{\ell^2}\}. \tag{5.8}$$

By examining the smallest eigenvalues of $\Xi_{h,\ell}^{-1}(\Xi_{h,\ell}^{-1})^*$ it can be shown that $\|\Xi_{h,\ell}\|_{\ell^2} \le 1 + O(|\zeta_{h,\ell}|)$, and hence it follows from Lemma 4.13 that

$$q_Q \approx \max_{\ell>\ell_*}\{|Q_{h,\ell}|\}.$$

The estimates of the double sweep operator for modes of $\ell > \ell_*$ are given in the following lemma resulting from (5.6) and (5.7).

Lemma 5.2 *Assume that $h < h_1$. For $\ell > \ell_*$, it holds that*

$$\|\boldsymbol{\gamma}_{h,>}^{\mathrm{L},m}\|_{\dot{H}_h^{-1/2}(\Gamma)} \le q_Q\|\boldsymbol{\gamma}_{h,>}^{\mathrm{R},m}\|_{\dot{H}_h^{-1/2}(\Gamma)},$$
$$\|\boldsymbol{\gamma}_{h,>}^{\mathrm{R},m+1}\|_{\dot{H}_h^{-1/2}(\Gamma)} \le q_Q^2\|\boldsymbol{\gamma}_{h,>}^{\mathrm{R},m}\|_{\dot{H}_h^{-1/2}(\Gamma)}.$$

5.3.3 Convergence

Here we combine all estimates in the above to have the contraction property of $\mathbb{B}_h \circ \mathbb{F}_h : G_h^{\mathrm{R}} \to G_h^{\mathrm{R}}$. We introduce contraction factors

$$\mathfrak{q} = \max\{q_Q, C_\delta(J-1)e^{-2\sigma_\mu \beta}\} \quad \text{and} \quad \mathfrak{r} = \max\{q_Q^2, C_\delta(J-1)e^{-2\sigma_\mu \beta}\},$$

where the maximal values are taken among all modes involved in traces of approximate solutions generated by the double sweep DDM.

Lemma 5.3 *Assume the conditions in Lemma 5.1. Then it holds that*

$$\|\boldsymbol{\gamma}_h^{\mathrm{L},m}\|_{\dot{H}_h^{-1/2}(\Gamma)} \le \mathfrak{q}\|\boldsymbol{\gamma}_h^{\mathrm{R},m}\|_{\dot{H}_h^{-1/2}(\Gamma)}, \tag{5.9}$$

$$\|\boldsymbol{\gamma}_h^{\mathrm{R},m+1}\|_{\dot{H}_h^{-1/2}(\Gamma)} \le \mathfrak{r}\|\boldsymbol{\gamma}_h^{\mathrm{R},m}\|_{\dot{H}_h^{-1/2}(\Gamma)}. \tag{5.10}$$

Proof The orthogonality of the basis leads us to

$$\|\boldsymbol{\gamma}_h^{\mathrm{R},m+1}\|^2_{\dot{H}_h^{-1/2}(\Gamma)} = \|\boldsymbol{\gamma}_{h,\le}^{\mathrm{R},m+1}\|^2_{\dot{H}_h^{-1/2}(\Gamma)} + \|\boldsymbol{\gamma}_{h,>}^{\mathrm{R},m+1}\|^2_{\dot{H}_h^{-1/2}(\Gamma)}.$$

By invoking Lemma 5.1 and Lemma 5.2, we have the contraction property (5.10) of the double sweep operator,

$$\|\boldsymbol{\gamma}_h^{\mathrm{R},m+1}\|^2_{\dot{H}_h^{-1/2}(\Gamma)} \le \mathfrak{r}^2(\|\boldsymbol{\gamma}_{h,\le}^{\mathrm{R},m}\|^2_{\dot{H}_h^{-1/2}(\Gamma)} + \|\boldsymbol{\gamma}_{h,>}^{\mathrm{R},m}\|^2_{\dot{H}_h^{-1/2}(\Gamma)}) = \mathfrak{r}^2\|\boldsymbol{\gamma}_h^{\mathrm{R},m}\|^2_{\dot{H}_h^{-1/2}(\Gamma)}.$$

The proof of (5.9) can be done in the same way and the proof is completed. □

Finally, we can establish the convergence of the discrete double sweep DDM in $H^1(\Omega)$.

Theorem 5.4 *Suppose that $f = 0$. Let $u_h^m \in V_\Omega^h$ be the m-th iterate of the double sweep DDM for any $u_h^0 \in V_\Omega^h$. Assume that $0 < h < h_1$ and c_{0h} is a positive constant satisfying $c_{0h}(J-1) < 1$ as well as $(2C_d c_{0h}^2 + C_d^2 c_{0h}^4)(J-1) < 1$. Let K_j^{R} for $j = 1, \dots, J-1$ be the set of all triangulations $\tau \in \mathcal{T}_{\Omega_j}$ that have at least one vertex on Γ_j and denote $\Omega^{\mathrm{R}} = \Omega \setminus (\cup_{j=1}^{J-1} \cup_{\tau \in K_j^{\mathrm{R}}} \bar{\tau})$. If the* PML *parameters are chosen such that $e^{-2\sigma_\mu \beta} < c_{0h}$, then it holds that*

$$\|u_h^m\|_{H^1(\Omega^{\mathrm{R}})} \lesssim (\mathfrak{r} + \mathfrak{q})\mathfrak{r}^{m-1}\|u_h^0\|_{H^1(\Omega)}.$$

Furthermore, we have the L^2-norm convergence in the whole domain Ω,

$$\|u_h^m\|_{L^2(\Omega)} \lesssim (\mathfrak{r} + \mathfrak{q})\mathfrak{r}^{m-1}\|u_h^0\|_{H^1(\Omega)}.$$

Proof Let $\tilde{\boldsymbol{u}}_h^m = (\tilde{u}_{h,1}^m, \dots, \tilde{u}_{h,J}^m) \in \prod_{j=1}^J V_{\Omega_j}^h$, where $\tilde{u}_{h,j}^m$ is the solution to the local problem in Ω_j with the boundary data $\gamma_{h,j}^{\mathrm{L},m-1}$ and $\gamma_{h,j}^{\mathrm{R},m}$, obtained during the backward sweep. By the same argument used for Theorem 3.8 with Lemma 3.4 and Lemma 3.6 replaced by Lemma 5.3, we can show that

$$\begin{aligned}\|\tilde{\boldsymbol{u}}_h^m\|_{\boldsymbol{V}} &\lesssim \|\boldsymbol{\gamma}_h^{\mathrm{R},m}\|_{\dot{H}_h^{-1/2}(\Gamma)} + \|\boldsymbol{\gamma}_h^{\mathrm{L},m-1}\|_{\dot{H}_h^{-1/2}(\Gamma)} \\ &\lesssim (\mathfrak{r} + \mathfrak{q})\|\boldsymbol{\gamma}_h^{\mathrm{R},m-1}\|_{\dot{H}_h^{-1/2}(\Gamma)} \lesssim (\mathfrak{r} + \mathfrak{q})\mathfrak{r}^{m-1}\|u_h^0\|_{H^1(\Omega)}.\end{aligned}$$

Here we used $\|\boldsymbol{\gamma}_h^{\mathrm{R},0}\|_{\dot{H}_h^{-1/2}(\Gamma)} \lesssim \|\tilde{\boldsymbol{u}}_h^0\|_{\boldsymbol{V}}$ and $\|\tilde{\boldsymbol{u}}_h^0\|_{\boldsymbol{V}} = \|u_h^0$
$|_{H^1(\Omega)}$ in the last inequality. Since $\tilde{u}_{h,J}^m = u_{h,J}^m$ and $v_{h,j}^m := \tilde{u}_{h,j}^m - u_{h,j}^m$ for $j = 1, \ldots, J-1$ is supported in $\cup_{\tau \in K_j^{\mathrm{R}}} \bar{\tau}$, it is obvious that

$$\|u_h^m\|_{H^1(\Omega^{\mathrm{R}})} \leq \|\tilde{\boldsymbol{u}}_h^m\|_{\boldsymbol{V}} \lesssim (\mathfrak{r} + \mathfrak{q})\mathfrak{r}^{m-1}\|u_h^0\|_{H^1(\Omega)}.$$

In addition, a simple computation gives

$$\|v_{h,j}^m\|_{L^2(\Omega_j)}^2 = \sum_{\tau \in K_j^{\mathrm{R}}} \|v_{h,j}^m\|_{L^2(\tau)}^2 = \sum_{e \in \mathcal{T}_{\Gamma_j}} \frac{h}{3}\|v_{h,j}^m\|_{L^2(e)}^2 = \frac{h}{3}\|v_{h,j}^m\|_{H^{1/2}(\Gamma_j)}^2.$$

Since the trace of $u_{h,j}^m$ on Γ_j coincides with that of $\tilde{u}_{h,j+1}^m$, it is obtained by a trace inequality that

$$\sum_{j=1}^{J-1} \|v_{h,j}^m\|_{H^{1/2}(\Gamma_j)}^2 \leq 2\sum_{j=1}^{J-1} \left(\|\tilde{u}_{h,j}^m\|_{H^{1/2}(\Gamma_j)}^2 + \|\tilde{u}_{h,j+1}^m\|_{H^{1/2}(\Gamma_j)}^2\right) \lesssim \|\tilde{\boldsymbol{u}}_h^m\|_{\boldsymbol{V}}^2.$$

Therefore combining the above estimates gives

$$\begin{aligned}\|u_h^m\|_{L^2(\Omega)}^2 &\leq \|\tilde{u}_{h,J}^m\|_{L^2(\Omega)}^2 + 2\sum_{j=1}^{J-1} \left(\|\tilde{u}_{h,j}^m\|_{L^2(\Omega_j)}^2 + \|v_{h,j}^m\|_{L^2(\Omega_j)}^2\right) \\ &\lesssim \|\tilde{\boldsymbol{u}}_h^m\|_{\boldsymbol{V}}^2 \lesssim (\mathfrak{r} + \mathfrak{q})^2\mathfrak{r}^{2(m-1)}\|u_h^0\|_{H^1(\Omega)}^2,\end{aligned}$$

which completes the proof. □

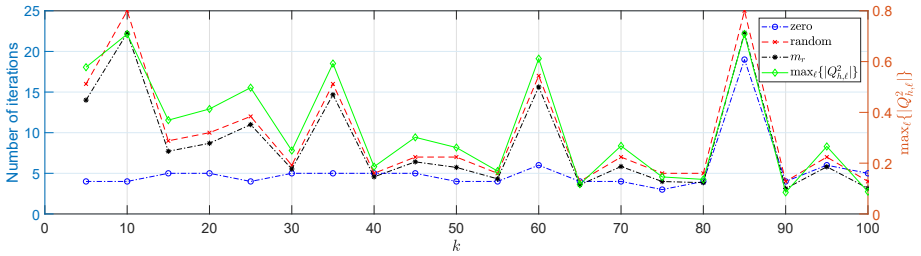
As a consequence of this theorem, since the contraction factor for the propagating modes depends linearly on the number of subdomains, the number of iteration increases only logarithmically with respect to the number of subdomains.

6 Numerical experiments

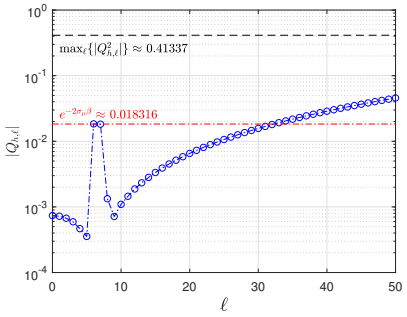
In this section, we provide numerical examples illustrating the convergence theory studied in the preceding section. The domain Ω in this section is set to be a rectangular one and the left-side boundary of Ω is assigned for PML and the homogeneous Neumann boundary condition are given on all other boundaries. The problem is discretized with the help of the finite element library deal.II [2] and the solution is approximated by using continuous piecewise bilinear finite elements. The double sweep DDM is applied to the finite element problem with the stopping criterion that residuals relative to the initial residual are less than 10^{-5}.

6.1 Dominant contraction factor

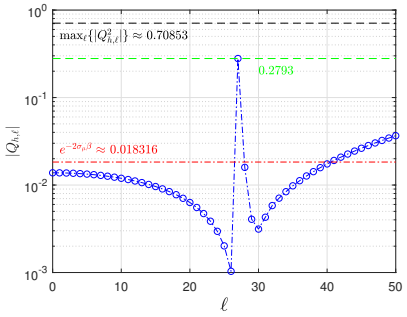
In the first experiment, we discuss which contraction factor between $C_\delta(J-1)e^{-2\sigma_\mu\beta}$ and q_Q^2 is dominant provided h is small enough. The domain Ω is the unit square $\Omega = (0,1) \times (0,1)$ and the finite element mesh nodes are deployed with $h = 1/400$ on the domain. We take somewhat large $\beta = 0.1$ (which results in large $N_p = 40$) in this example to avoid excessively slow convergence when q_Q^2 is dominant. For



(a) Number of iterations vs. k



(b) Reflection coefficients for $k = 20$



(c) Reflection coefficients for $k = 85$

Fig. 3: $it_\#$ vs. k when $J = 5$, $\sigma_\mu = 20$ and $\beta = 0.1$ with $e^{-2\sigma_\mu\beta} \approx 0.0183 < \max_\ell\{|Q^2_{h,\ell}|\} \approx q_Q^2$. In (a), the blue dash-dot curve : $it_\#$ for with zero initial iterates; the red dash curve : $it_\#$ for random initial iterates.

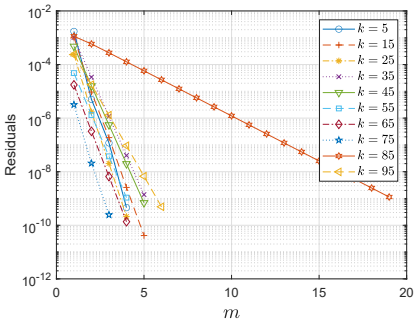
both cases, the PML parameter $\sigma_\mu = 20$ is taken so that the continuous reflection coefficient is bounded by $e^{-2\sigma_\mu\beta} \approx 0.0183$.

We consider a source problem with a compactly supported L^2 source function f defined by

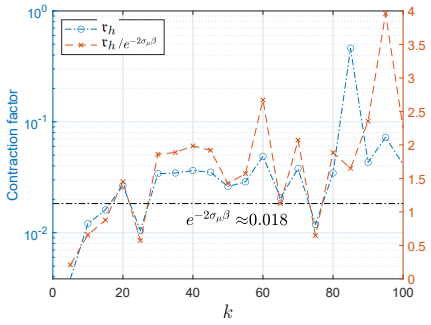
$$f(x,y) = \begin{cases} 1 & \text{if } \|(x,y) - (0.1, 0.2)\|_{\ell^2} \le 0.05, \\ 0 & \text{otherwise} \end{cases} \tag{6.1}$$

and the double sweep DDM with $J = 5$ starts with zero or random initial iterates. Fig. 3 (a) reports the number of iterations, denoted by $it_\#$, of the double sweep DDM vs. wavenumbers from 5 to 100. The blue dash-dot curve represents $it_\#$ of the double sweep DDM starting with zero initial iterates, and it shows that at most five iterations are enough to obtain the desired approximate solutions for all wavenumbers except for $k = 85$ and it appears that the number of iterations does not depend on wavenumbers. According to Fig. 3 (c), it looks that the peak at $k = 85$ is caused by the fact that the mesh size $h = 1/400$ is not small enough for $Q_{h,\ell}$ of near-cutoff modes for $k = 85$ to drop down to the level of $e^{-2\sigma_\mu\beta}$ unlike cases for other k shown in Fig. 3 (b), for example $k = 20$.

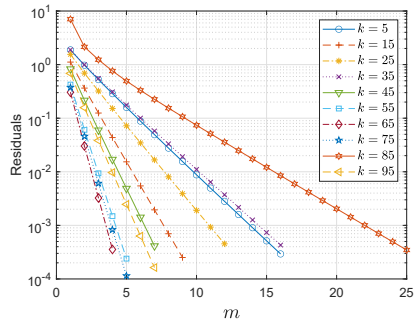
The residuals of the m-th iterates are evaluated and they are given in Fig. 4 (a). We also estimate the numerical contraction factors $\mathfrak{r}_h$ (the ratio of two consecutive residuals) shown in Fig. 4 (b), where the left y-axis represents the contraction factors and the right y-axis stands for the values $\mathfrak{r}_h/e^{-2\sigma_\mu\beta}$, constant multiples of the



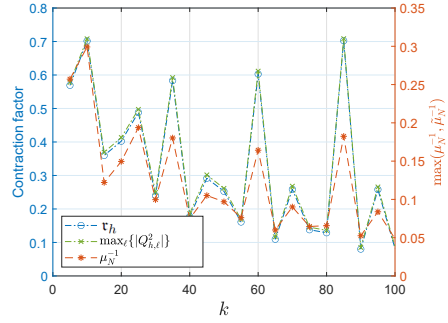
(a) Residuals with zero initial iterates



(b) $\mathfrak{r}_h$ with zero initial iterates



(c) Residuals with random initial iterates



(d) $\mathfrak{r}_h$ with random initial iterates

Fig. 4: Residuals of the m-th iterate and numerical contraction factor $\mathfrak{r}_h$ obtained by the ratio of two consecutive residuals.

maximal continuous reflection coefficient. They exhibit that the sequences generated by the double sweep DDM converge linearly with contraction factor which are constant multiples of the maximal continuous reflection coefficient ranged between 0.2 and 4. For $k = 85$, the constant multiple is computed with respect to the peak 0.2793 in Fig. 3 (c) instead of $e^{-2\sigma_\mu\beta}$. Thus it can be concluded that if support of the source function f is located away from interfaces of subdomains and the double sweep DDM starts with a zero initial iterates, then the convergence in these examples is controlled by the maximal continuous reflection coefficients since the residuals include only propagating modes and relatively slowly decaying evanescent modes in their traces on the interfaces of subdomains.

On the other hand, the red dash curve in Fig. 3 (a) shows $it_\#$ of the double sweep DDM starting with random initial iterates. Convergence of this case is slow compared with that of the double sweep DDM with zero initial iterates but they still converge linearly (see Fig. 4 (c)). In case that $\mathfrak{q} < C\mathfrak{r}$, Theorem 5.4 shows that the number of iterations $it_\#$ required for residuals relative to the initial residual to be within given tolerance τ can be estimated by $m_r := \frac{\ln(\tau/\|r_0\|_{L^2})}{\ln \mathfrak{r}}$, where r_0 stands for the initial residual. Interestingly, the shape of the plot of $it_\#$ looks similar to that of m_r given by the black dash-dot curve in Fig. 3 (a) as a function

of wavenumbers when $\mathfrak{r} = q_Q^2$ and $\tau = 10^{-5}$. Thus, by observing that $e^{-2\sigma_\mu \beta} \approx 0.0183 < \max_\ell\{|Q_{h,\ell}^2|\}$ from the solid green curve in Fig. 3 (a), the slow convergence of the double sweep DDM is explained by the existence of fast decaying evanescent modes in random initial iterates whose discrete reflection coefficients are dominant. This is also supported by Fig. 4 (d) demonstrating that the numerical contraction factors coincide with the actual contraction factors governed by $\max_\ell\{|Q_{h,\ell}^2|\} \approx q_Q^2$ in theory. Another interesting observation is that the performance of the double sweep DDM with random initial iterates is different wavenumber-by-wavenumber as opposed to that starting with a zero initial iterate. In fact, the discrete maximal reflection coefficient depends on σ_0 as shown in the asymptotic formula (4.44). That is, the performance of the method depends on the relative position of k with respect to the distribution of the cross-sectional eigenvalues. In particular, when there exist near-cutoff modes corresponding to $\mu_N \ll 1$ or $\tilde{\mu}_{N+1} \ll 1$, which are known to be troublesome for PML, the performance of the method is deteriorated and this is supported by the plot of $\max\{\mu_N^{-1}, \tilde{\mu}_{N+1}^{-1}\}$ given in Fig. 4 (d) showing the same shape as that of the maximal reflection coefficients.

6.2 Influence of the transmission condition based on PML

The next experiments are to examine the influence of the transmission conditions depending on two PML parameters σ_μ and β. We take f given by (6.1) in the domain $\Omega = (0,1) \times (0,1)$ and set $J = 5$ and $h = 1/400$.

1 Case I: varying σ_μ with $\beta = 0.1$ fixed.

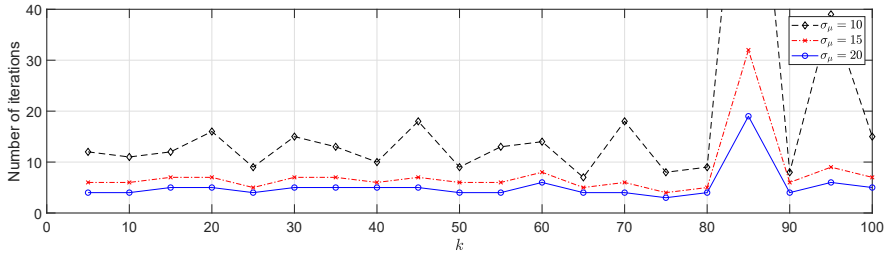
For the first case, we examine the performance of the method with different values of σ_μ with $\beta = 0.1$ fixed. When the double sweep DDM is fed a zero vector or a random vector for an initial iterate, we obtain the results presented in Fig. 5 (a) and (b), respectively.

In case of the double sweep DDM with zero initial iterates for $\sigma_\mu = 10$, 15 and 20, we see that $it_\#$ decreases monotonically with increasing σ_μ. In fact, when h is small enough and $\mathfrak{r} = C_\delta(J-1)e^{-2\sigma_\mu\beta}$, the contraction factor decreases as $\sigma_\mu\beta$ increases.

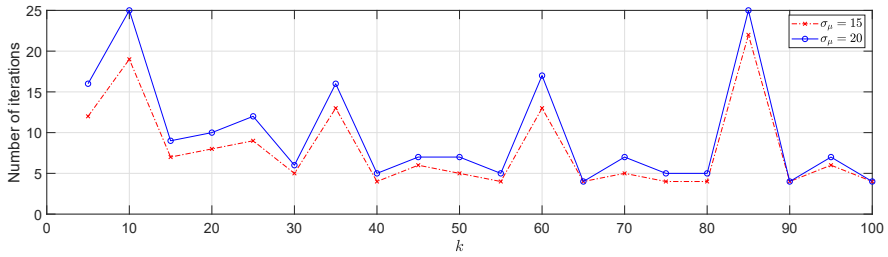
Next, we compare the performance of the double sweep DDM starting with random vectors when $\sigma_\mu = 15$ and 20 with $\beta = 0.1$ fixed. When a random vector is chosen for the first iterate, q_Q^2 is a dominant factor as seen in the experiments of Subsection 6.1. Interestingly, Fig. 5(c) and (d) show that $|Q_{h,\ell}|$ near the cutoff index N becomes smaller as we take larger σ_μ as expected, however the values of $|Q_{h,\ell}|$ are arranged in the reverse order far away from the cutoff index. This can be explained by examining the asymptotic formula (4.44) of the discrete maximal reflection coefficient. Thus, if q_Q^2 is dominant over $C_\delta(J-1)e^{-2\sigma_\mu\beta}$ for each σ_μ, then $it_\#$ decreases with decreasing σ_μ, which can be observed for most wavenumbers except for $k = 65$, 90 and 100 in Fig. 5 (b).

2 Case II: varying β with $\sigma_\mu = 10$ fixed.

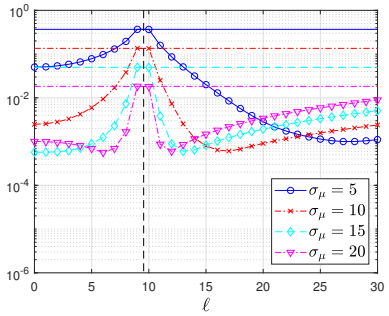
For the second case, we examine the performance of the method with different values of β with $\sigma_\mu = 10$ fixed. The results obtained with zero and random initial iterates are presented in Fig. 6 (a) and (b), respectively. When the double sweep DDM starts with a zero vector for $\beta = 0.1$, 0.2, 0.3 and so the contraction factor $\mathfrak{r}$ is governed by the term $C_\delta(J-1)e^{-2\sigma_\mu\beta}$, $it_\#$ decreases monotonically with increasing β as shown in Fig. 6 (a). When the double sweep DDM starts with a



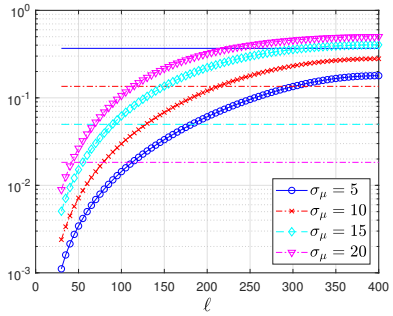
(a) Starting with a zero initial iterate



(b) Starting with a random initial iterate



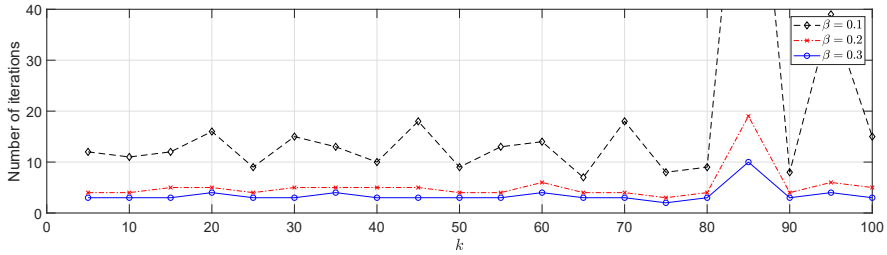
(c) $Q_{h,\ell}$ for $0 \le \ell \le 30$.



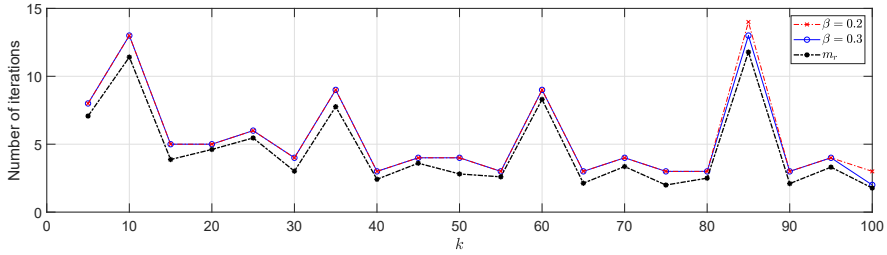
(d) $Q_{h,\ell}$ for $30 \le \ell \le 400$.

Fig. 5: $it_\#$ vs. σ_μ with $\beta = 0.1$ and $J = 5$. In (c) and (d), plots of $|Q_{h,\ell}|$ for $k = 30$. The horizontal lines in the plots represent the maximal continuous reflection coefficients $e^{-2\sigma_\mu\beta}$ for the PML parameters σ_μ and β corresponding to the same color and the same line style.

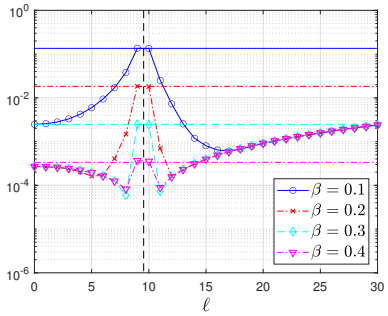
random vector, however, q_Q^2 can be dominant over $C_\delta(J-1)e^{-2\sigma_\mu\beta}$. According to Fig. 6 (c) and (d), the maximal discrete reflection coefficients do not depend on β, which is justified by the asymptotic formula (4.44) independent of β for sufficiently large β, so that $it_\#$ can be constant for all sufficiently large β if the contraction factor $\mathfrak{r}$ is controlled by q_Q^2. These results can be observed in Fig. 6 (b), where the red dash-dot curve of $\beta = 0.2$ coincides with the blue solid curve of $\beta = 0.3$ for all k except $k = 85$ and 100.



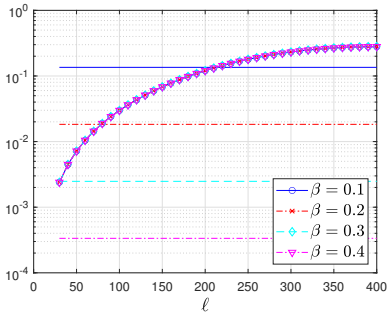
(a) Starting with a zero initial iterate



(b) Starting with a random initial iterate



(c) $Q_{h,\ell}$ for $0 \le \ell \le 30$



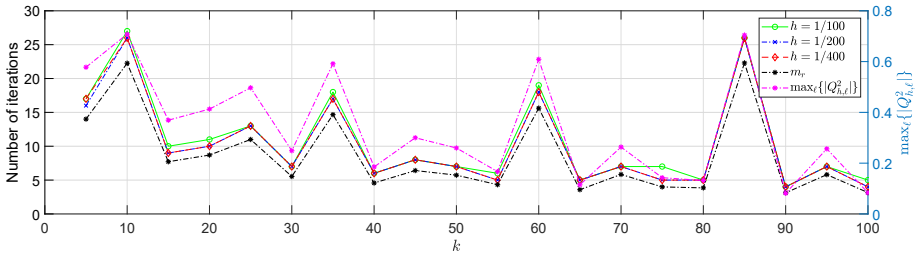
(d) $Q_{h,\ell}$ for $30 \le \ell \le 400$

Fig. 6: $it_{\#}$ vs. β with $\sigma_\mu = 10$ and $J = 5$. In (c) and (d), plots of $Q_{h,\ell}$ for $k = 30$. The horizontal lines in the plots represent the maximal continuous reflection coefficients $e^{-2\sigma_\mu \beta}$ for the PML parameters σ_μ and β corresponding to the same color and the same line style.

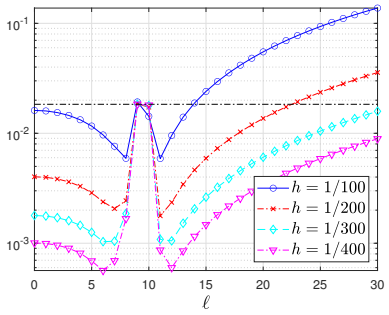
6.3 Dependence on h

The next experiment exhibits the behavior of the double sweep DDM with respect to h. Here we consider two examples. The first one is for h-independent performance of the double sweep DDM. We take $\Omega = (0,1) \times (0,1)$ and set $\sigma_\mu = 20$, $\beta = 0.1$. The domain Ω is decomposed into 5 subdomains, i.e. $J = 5$ and so $H = 0.2$, and the finite element method with $h = 1/100$, $1/200$ and $1/400$ is applied.

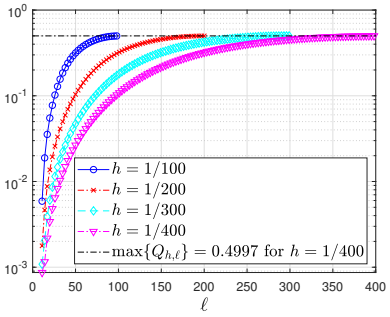
As for the behavior of $Q_{h,\ell}$ with respect to h, we see Fig. 7 (b) and (c) showing that the discrete reflection coefficients for each ℓ get smaller as h decreases, however



(a) Starting with a random initial iterate



(b) $Q_{h,\ell}$ for $0 \le \ell \le 30$

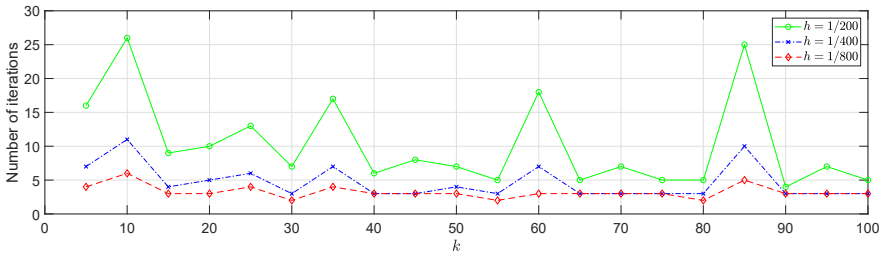


(c) $Q_{h,\ell}$ for $30 \le \ell \le 400$

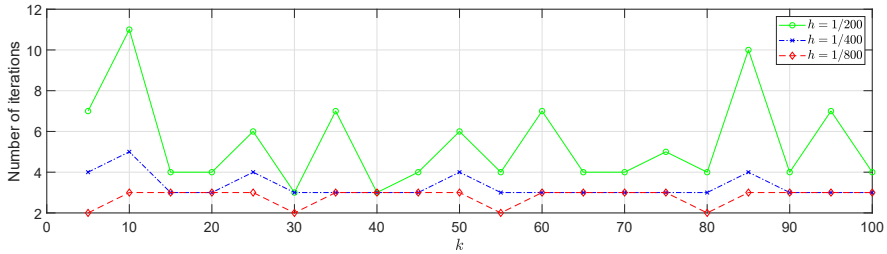
Fig. 7: $it_{\#}$ vs. h for random initial iterates with $\sigma_\mu = 20$, $\beta = 0.1$ and $J = 5$. In (b) and (c), plots of $Q_{h,\ell}$ for $k = 30$. The horizontal line in the plots represents the maximal continuous reflection coefficients $e^{-2\sigma_\mu \beta}$.

the maximal discrete reflection coefficients for different values of h with σ_μ and β fixed do not have noticeable difference. This is because the upper bound of x in the right hand side of (4.44) is independent of h. The double sweep DDM with random initial iterates produces the results in Fig. 7 (a), showing that the convergence of the double sweep DDM is governed by $q_Q^2 \approx \max_\ell\{|Q_{h,\ell}^2|\}$. Since $\max_\ell\{|Q_{h,\ell}^2|\}$ do not have any significant change for different $h = 1/100$, $1/200$ and $1/400$, only one plot for $h = 1/400$ is reported with its corresponding reference number of iterations m_r in Fig. 7.

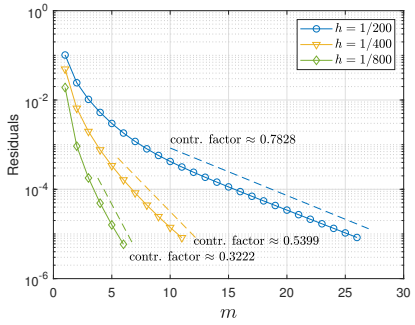
In contrast, the convergence of the double sweep DDM may depend on h if approximate solutions includes only a partial set of modes such as propagating modes and slowly decaying evanescent modes. As an example, we take the domain $\Omega = (0, 0.2) \times (0, 1)$ and decompose it into 5 subdomains so that $J = 5$ and $H = 0.04$ and use $h = 1/200$, $1/400$ and $1/800$ for finite element approximations. Assume that f is a point source located at $(0.05, 0.5)$. By using PML with $\sigma_\mu = 30$ and $\beta = 0.1$, we obtain $it_{\#}$ shown in Fig. 8 (a) revealing that the performance of the double sweep DDM is improved as h is getting smaller. In order to examine the convergence with respect to h in more detail, we estimate the convergence factors for $k = 10$ in Fig. 8 (c), exhibiting that numerical contraction factors are approximately 0.7828, 0.5399 and 0.3222 for each h respectively, compared with $\max_\ell\{|Q_{h,\ell}^2|\} \approx 0.7936$. Since reflection coefficients for propagating modes



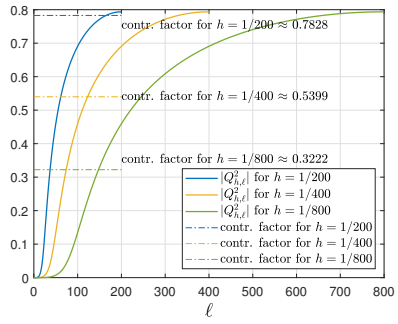
(a) f is a point source at $(0.05, 0.5)$



(b) f is a point source at $(0.02, 0.5)$



(c) Residuals



(d) Estimated number of modes

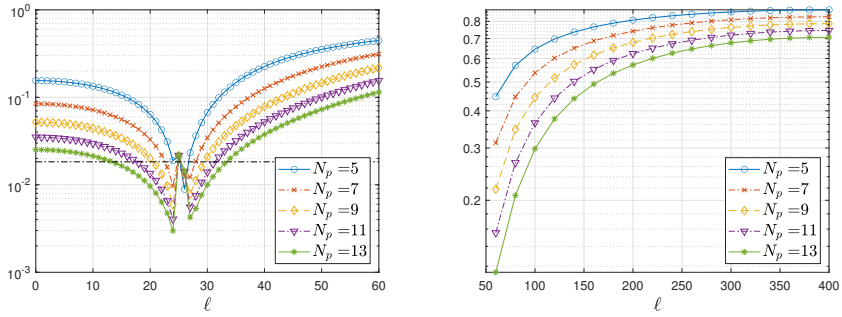
Fig. 8: $it_{\#}$ vs. h for zero initial iterates with $\sigma_\mu = 30$, $\beta = 0.1$, $J = 5$ and $H = 0.04$. In (c) and (d), residuals and estimated number of modes in residuals for $k = 10$.

are less than 10^{-2} for all h and they are significantly smaller than the numerical contraction factors calculated as above, we infer that the contraction factor $\mathfrak{r}$ is determined by q_Q^2. In addition, according to Fig. 8 (d), we find that there are roughly $140 \sim 180$ modes involved in traces of approximate solutions on interfaces.

At last, it is worth pointing out that the modes involved in traces of approximate solutions on interfaces are related with the distance from the source to interfaces. If a point source is located farther from interfaces of subdomains, then more evanescent modes emitting from the source can diminish on interfaces and the contraction factors become smaller, which results in the smaller numbers of iterations. For example, we consider a point source located at $(0.02, 0.5)$. In this case the distance from the source to the interfaces is 0.02, which is greater than

J \ N_p	5	6	7	8	9	10	11	12	13	14	15
4	13	10	9	8	7	6	6	5	5	5	4
8	18	15	13	11	10	9	8	7	7	6	6
16		25	14	12	11	10	9	8	8	7	7
5	16	13	11	10	9	8	7	7	6	6	5
10	44	16	13	12	10	9	8	8	7	7	6
20		22	15	13	11	10	9	9	8	8	7

Table 1: $it_{\#}$ with zero initial iterates for $k = 80$, $h = 1/400$. Here σ_μ is chosen such that $e^{-2\sigma_\mu \beta} \approx 0.0183$ for varying N_p.



(a) $Q_{h,\ell}$ for $0 \le \ell \le 60$ (b) $Q_{h,\ell}$ for $60 \le \ell \le 400$

Fig. 9: Discrete reflection coefficients $Q_{h,\ell}$ for $k = 80$ with $h = 1/400, \sigma_\mu \beta = 2$.

the distance ($= 0.01$) from the source at $(0.05, 0.5)$. We see that the numbers of iterations for the case of the point source at $(0.02, 0.5)$ given in Fig. 8 (b) are smaller than those for the point source at $(0.05, 0.5)$.

6.4 Dependence on J, the number of subdomains

In this subsection we conduct experiments to see how the double sweep DDM depends on J. To do this, we take the square domain $\Omega = (0, 1) \times (0, 1)$ and decompose Ω into uniform quadrilateral finite elements with $h = 1/400$. Through this test, we also investigate how large the PML width $\beta = hN_p$ is required to keep $it_{\#}$ constant as J increases. The PML strength σ_μ is chosen such that the reflection coefficient $e^{-2\sigma_\mu \beta}$ is constant for varying β. For $k = 80$, the test results of the double sweep DDM with zero initial iterates for $J = 4, 5, 8, 10, 16$ and 20, are shown in Table 1. As described in Theorem 5.4 when $C_\delta (J-1) e^{-2\sigma_\mu \beta}$ is the dominant reflection coefficient, it is observed in Table 1 that

$$it_{\#} = O(\ln(J)). \tag{6.2}$$

for each N_p. Since it is of importance to understand how fast N_p needs to increase to keep $it_{\#}$ constant with growing J in the practical use of the double sweep DDM, we also examine N_p with the same number of iterations as J increases, for instance,

we consider $it_{\#}$ in the cells with gradual color in Table 1. They demonstrate that N_p grows logarithmically with respect to the number of subdomains,

$$N_p = O(\ln(J)), \tag{6.3}$$

which has been already noticed in [36] of studying the double sweep DDM with PML of quadratic stretching functions in the open space $\mathbb{R}^2$. In fact, this result (6.3) can be explained in terms of the discrete reflection coefficients as shown in Fig. 9. It reveals that $Q_{h,\ell}$ decreases exponentially for each ℓ as N_p increases linearly in the above the black dash-dot line representing $e^{-2\sigma_\mu\beta}$. It means that

$$N_p = O(\ln(1/\mathfrak{r})) = O(it_{\#}^{-1}). \tag{6.4}$$

This result can be found in each row of Table 1 by comparing $it_{\#}$ for the pairs of N_p, for example, $(5, 10)$, $(6, 12)$ and $(7, 14)$. Finally, it follows from (6.2) and (6.4) that $it_{\#}$ is proportional to $\ln(J)$ and $1/N_p$, which leads to (6.3) to keep $it_{\#}$ constant.

6.5 Experiments with respect to N_p with $\sigma_\mu\beta$ fixed

Next, we present some experiments in the square domain $\Omega = (0, 1) \times (0, 1)$ of the double sweep DDM of PML with respect to N_p with $\sigma_\mu\beta$ fixed. The earlier research of the double sweeping preconditioners based on PML with quadratic stretching functions in [12,35,36] studied computational costs of the method with respect to N_p since it is preferable to minimize the computational cost by taking as small N_p as possible. Assuming that n is the number of the grid points along each axis of the physical domain Ω, [12] shows that the double sweep preconditioning technique requires $O(N_p^2 n^2)$ and illustrates numerical examples with $N_p = 10$ for its efficiency. Also, in [35] smaller N_p such as 5 or 6 is used in the domain decomposition framework and the double sweep DDM with $N_p = 3, 4$ or 5 in [36] is successfully used for coarse grid solvers in the multi-level framework. When the domain Ω is decomposed into J subdomains of equal size, the computational costs for one sweep are of order $O((\frac{n}{J} + 2N_p)^2 nJ)$. From (6.3), it can be seen that N_p for the minimal computational costs grows of order $O(\ln(n))$. Indeed, for the asymptotic analysis, we just replace J with e^{N_p} in the asymptotic cost rate $(\frac{n}{J}+2N_p)^2 nJ$ and find that the minimal cost occurs when N_p satisfies $(2+N_p)e^{N_p} \approx n$, which shows

$$N_p = O(\ln(n)) \text{ and } J = O(n) \tag{6.5}$$

for the same number of iterations. This result can be observed in Table 2 with σ_μ and $\beta = hN_p$ chosen so that the maximal reflection coefficients remain the same, $e^{-2\sigma_\mu\beta} \approx 0.0183$, for varying N_p. Therefore, it turns out that the computational cost is of order $O(N_p^2 n^2)$, which is the same order of the method studied in [12].

We further perform experiments of the double sweep DDM for PML with quadratic stretching functions defined by

$$\tilde{x} = \begin{cases} \frac{\sigma_0}{\beta^3} \int_0^x t^2 dt & \text{for } x > 0, \\ x & \text{for } x \le 0 \end{cases}$$

for the absorbing boundary condition and its analogues for transmission conditions with σ_0 as in Subsection 2.2 and compare them with those obtained with constant

h \ N_p	5	6	7	8	9	10	11	12	13	14
1/200	10	7	6	5	5	4	4	4	4	4
1/400	11	8	6	6	5	5	4	4	4	4
1/800	12	8	7	6	5	5	5	4	4	4

Table 2: $it_{\#}$ with zero initial iterates for $k = 80$, $J = 5$. Here σ_μ is chosen so that $e^{-2\sigma_\mu \beta} \approx 0.0183$ for varying N_p.

k \ N_p	piecewise constant				piecewise quadratic			
	5	10	20	40	5	10	20	40
20	6	5	5	5	6	5	5	5
55	10	5	4	4	6	4	4	4
80	11	5	4	4	5	4	4	4
95	15	6	6	6	7	6	6	6

Table 3: $it_{\#}$ with zero initial iterates for $J = 5$ and $h = 1/400$ in waveguides.

k \ N_p	piecewise constant				piecewise quadratic			
	5	10	20	40	5	10	20	40
20	75	38	19	10	8	4	4	3
55	37	19	10	5	6	3	3	3
80	32	16	8	5	6	3	3	3
95	48	24	12	7	8	4	4	4

Table 4: $it_{\#}$ with random initial iterates for $J = 5$ and $h = 1/400$ in waveguides.

stretching functions of this paper. We take a constant $\sigma_\mu = 6$ for quadratic stretching functions, because the stretching function is normalized by the PML width, so that the reflection coefficient is equal to that for PML with constant stretching functions, $e^{-2\sigma_\mu/3} \approx 0.0183$. Numbers of iterations of the double sweep DDM for $J = 5$, $h = 1/400$ and for several wavenumbers $k = 20, 55, 80, 95$ are reported in Table 3 and Table 4. As discussed in the previous subsection, it can be seen that $it_{\#}$ decreases with growing N_p. An important observation is that the performance of PML with quadratic stretching functions is better in particular, when the double sweep DDM starts with random initial iterates. It appears that fast decaying evanescent modes can be transmitted well by PML with quadratic stretching functions as opposed to PML with constant stretching functions.

We also apply the double sweep DDM to the Helmholtz equation in the open space $\mathbb{R}^2$ with only imaginary parts being stretched by PML. The results are presented in Table 5 and Table 6. It shows that PML of $N_p = 5$ can provide good transmission conditions for both cases of constant and quadratic stretching functions, when initial iterates are zero. However $it_{\#}$ increases drastically with decreasing N_p for the double sweep DDM with PML of constant stretching functions starting with random initial iterates. This phenomenon is more noticeable in the lower frequency regime. It seems that the solutions of low frequency in the open space correspond to near-cutoff modes of large wavelength in waveguides, for which large σ_0 is required for sufficient absorption and it makes in turn discrete reflection coefficients large.

k \ N_p	piecewise constant				piecewise quadratic			
	5	10	20	40	5	10	20	40
20	5	4	4	3	5	4	4	3
55	6	4	4	3	5	4	4	3
80	5	4	3	3	5	4	3	3
95	5	4	3	3	5	4	3	3

Table 5: $it_{\#}$ with zero initial iterates for $J = 5$ and $h = 1/400$ in $\mathbb{R}^2$.

k \ N_p	piecewise constant				piecewise quadratic			
	5	10	20	40	5	10	20	40
20	95	26	8	4	4	3	3	3
55	15	5	3	3	5	4	3	3
80	8	4	3	3	5	4	3	3
95	6	4	3	3	5	4	3	3

Table 6: $it_{\#}$ with random initial iterates for $J = 5$ and $h = 1/400$ in $\mathbb{R}^2$.

7 Appendix

In this section, we provide the proof of Lemma 4.9.

Proof (Proof of Lemma 4.9) We will estimate

$$\chi(w)^{2N_p} = \left((1 + iw + \mathcal{R}_\chi(w))^{\frac{1}{iw+\mathcal{R}_\chi(w)}}\right)^{\frac{2\beta}{h}(iw+\mathcal{R}_\chi(w))} := \mathfrak{a}^{\mathfrak{w}}$$

where

$$\mathfrak{a} = (1+iw+\mathcal{R}_\chi(w))^{\frac{1}{iw+\mathcal{R}_\chi(w)}} \text{ and } \mathfrak{w} = \frac{2\beta}{h}(iw + \mathcal{R}_\chi(w)) = 2\mu_{h,\ell}\sigma_0\beta\left(i + \frac{\mathcal{R}_\chi(w)}{w}\right).$$

Since $\mathfrak{a} \to e$ and $\mathcal{R}_\chi(w)/w \to 0$ as $w \to 0$, for any $\epsilon > 0$ we can take a positive constant δ small enough so that

$$\sigma_r \frac{|\arg(\mathfrak{a})|}{\ln|\mathfrak{a}|} < \sigma_i\epsilon \quad \text{and} \quad \left|\sigma_0 \frac{\mathcal{R}_\chi(w)}{w}\right| < \epsilon \min\{\sigma_r, \sigma_i\}. \tag{7.1}$$

for $|w| < \delta$.

Noting $|\mathfrak{a}^{\mathfrak{w}}| = e^{\Re(\mathfrak{w})\ln(|\mathfrak{a}|) - \Im(\mathfrak{w})\arg(\mathfrak{a})}$, we need to estimate $\Re(\mathfrak{w})$ and $\Im(\mathfrak{w})$. First, for $\mu_{h,\ell}^2 > 0$ by using the second inequality of (7.1) we have that

$$\Re(\mathfrak{w}) < -2\mu_{h,\ell}\sigma_i\beta(1-\epsilon) \quad \text{and} \quad |\Im(\mathfrak{w})| < 2\mu_{h,\ell}\sigma_r\beta(1+\epsilon).$$

The first inequality of (7.1) with the above estimates leads us to

$$\begin{aligned}\Re(\mathfrak{w})\ln(|\mathfrak{a}|) - \Im(\mathfrak{w})\arg(\mathfrak{a}) &< -2\mu_{h,\ell}\sigma_i\beta\ln(|\mathfrak{a}|)((1-\epsilon) - \epsilon(1+\epsilon)) \\ &< -2\mu_{h,\ell}\sigma_i\beta\ln(|\mathfrak{a}|)(1-3\epsilon).\end{aligned}$$

Due to the convergence of finite element eigenvalue approximations, there is $0 < \epsilon_h < \epsilon_*$ for each $0 < h \le h_0$ such that $\mu_{h,\ell} > \mu_{\min} - \epsilon_h$ for all ℓ and $\epsilon_h \to 0$ as $h \to 0$. Thus we can further show that for $0 < h \le h_0$

$$\Re(\mathfrak{w})\ln(|\mathfrak{a}|) - \Im(\mathfrak{w})\arg(\mathfrak{a}) < -2(1-3\epsilon)\beta\ln(|\mathfrak{a}|)(\sigma_\mu - \epsilon_h\sigma_i)$$

In the same way, one can show that for $\mu_{h,\ell}^2 < 0$

$$\Re(\mathfrak{w}) \ln(|\mathfrak{a}|) - \Im(\mathfrak{w}) \arg(\mathfrak{a}) < -2(1-3\epsilon)\beta \ln(|\mathfrak{a}|)(\sigma_\mu - \epsilon_h \sigma_r).$$

Thus, we have

$$\lim_{w\to 0} |\chi(w)|^{2N_p} \leq e^{2\epsilon_h(1-3\epsilon)\sigma_M \beta} e^{-2(1-3\epsilon)\sigma_\mu \beta}$$

where $\sigma_M = \max\{\sigma_r, \sigma_i\}$. Since ϵ can be arbitrarily small, it holds that

$$\lim_{w\to 0} |\chi(w)|^{2N_p} \leq e^{2\epsilon_h \sigma_M \beta} e^{-2\sigma_\mu \beta}.$$

From the fact that $e^{\epsilon_h \sigma_M \beta} \to 1$ as $h \to 0$, it then follows that for any $\epsilon > 0$ there exist $0 < \delta_0$ and $0 < \hat{h}_1 \leq h_0$ such that if $|w| \leq \delta_0$ and $0 < h \leq \hat{h}_1$, then

$$|\chi(w)|^{2N_p} \leq (1+\epsilon) e^{-2\sigma_\mu \beta}. \tag{7.2}$$

Next, we will prove that (7.2) still holds for $|w| > \delta_0$ and for sufficiently small h. To this end, we write $w = h\mu_{h,\ell}\sigma_0 = re^{i\theta}$, where $r = |w|$ and $\theta = \arg(w)$ with $0 < \theta < \pi$. Let $\widetilde{\chi}_\theta(r) = \chi(w)$ as a function of r. Noting that $|\widetilde{\chi}_\theta(r)|^2 = 1 - 2r\sin(\theta) + O(r^2)$ resulting from the asymptotic behavior of χ in (4.42), it is revealed that $|\widetilde{\chi}_\theta(r)|$ is a decreasing function near the origin. By taking into account the fact $\lim_{r\to\infty} |\widetilde{\chi}_\theta(r)| = 2 - \sqrt{3} < 1$, we can choose $0 < \delta_1 < \delta_0$ small enough so that

$$|\widetilde{\chi}_\theta(\delta_1)| = \max_{r \geq \delta_1} |\widetilde{\chi}_\theta(r)|. \tag{7.3}$$

Let $h_1 \leq \hat{h}_1$ be a positive constant such that $\mu_{\min} h_1 |\sigma_0| < \delta_1$. Now, it suffices to prove (7.2) for $h \leq h_1$ and $|w| \geq \delta_1$. For $|w| \geq \delta_1$, let $\hat{w} = \delta_1 w/|w|$, which can be written as $\hat{w} = h\hat{\mu}_{h,\ell}\sigma_0$ with $\hat{\mu}_{h,\ell} = \mu_{h,\ell}\delta_1/|w|$. Then $\hat{\mu}_{h,\ell}$ satisfies $\mu_{\min} < |\hat{\mu}_{h,\ell}|$ as

$$\mu_{\min} h \leq \mu_{\min} h_1 < \frac{\delta_1}{|\sigma_0|} = \frac{|\hat{w}|}{|\sigma_0|} = |\hat{\mu}_{h,\ell}| h. \tag{7.4}$$

Since $|\hat{w}| = \delta_1 \leq |w|$, (7.3) gives $|\chi(w)| \leq |\chi(\hat{w})|$, which in turn together with (7.2) and the fact that $|\hat{w}| < \delta_0$ and $|\hat{\mu}_{h,\ell}| > \mu_{\min} - \epsilon_h$ obtained from (7.4) shows that

$$|\chi(w)|^{2N_p} \leq |\chi(\hat{w})|^{2N_p} \leq (1+\epsilon) e^{-2\sigma_\mu \beta},$$

and the proof is completed. □

References

1. M. Arioli and D. Loghin. Discrete interpolation norms with applications. *SIAM J. Numer. Anal.*, 47(4):2924–2951, 2009.
2. W. Bangerth, R. Hartmann, and G. Kanschat. deal.II—a general-purpose object-oriented finite element library. *ACM Trans. Math. Software*, 33(4):24, 2007.
3. A. Bermúdez, L. Hervella-Nieto, A. Prieto, and R. Rodríguez. An optimal perfectly matched layer with unbounded absorbing function for time-harmonic acoustic scattering problems. *J. Comput. Phys.*, 223(2):469–488, 2007.
4. J. H. Bramble. *Multigrid methods*, volume 294 of *Pitman Research Notes in Mathematics Series*. Longman Scientific & Technical, Harlow; copublished in the United States with John Wiley & Sons, Inc., New York, 1993.
5. J. H. Bramble and J. E. Pasciak. Analysis of a Cartesian PML approximation to acoustic scattering problems in $\mathbb{R}^2$ and $\mathbb{R}^3$. *J. Comput. Appl. Math.*, 247:209–230, 2013.

6. A. Brandt and I. Livshits. Wave-ray multigrid method for standing wave equations. *Electron. Trans. Numer. Anal.*, 6(Dec.):162–181, 1997. Special issue on multilevel methods (Copper Mountain, CO, 1997).
7. Z Chen and X Liu. An adaptive perfectly matched layer technique for time-harmonic scattering problems. *SIAM J. Numer. Anal.*, 43(2):645–671, 2005.
8. Z. Chen and X. Xiang. A source transfer domain decomposition method for Helmholtz equations in unbounded domain. *SIAM J. Numer. Anal.*, 51(4):2331–2356, 2013.
9. S. Cools and W. Vanroose. Local Fourier analysis of the complex shifted Laplacian preconditioner for Helmholtz problems. *Numer. Linear Algebra Appl.*, 20(4):575–597, 2013.
10. B. Després. *Méthodes de décomposition de domaine pour les problèmes de propagation d'ondes en régimes harmoniques*. PhD thesis, Université Dauphine - Parix, 1991.
11. V. Druskin, S. Güttel, and L. Knizhnerman. Near-optimal perfectly matched layers for indefinite Helmholtz problems. *SIAM Rev.*, 58(1):90–116, 2016.
12. B. Engquist and L. Ying. Sweeping preconditioner for the Helmholtz equation: moving perfectly matched layers. *Multiscale Model. Simul.*, 9(2):686–710, 2011.
13. Y. A. Erlangga, C. W. Oosterlee, and C. Vuik. A novel multigrid based preconditioner for heterogeneous Helmholtz problems. *SIAM J. Sci. Comput.*, 27(4):1471–1492, 2006.
14. Y. A. Erlangga, C. Vuik, and C. W. Oosterlee. On a class of preconditioners for solving the Helmholtz equation. *Appl. Numer. Math.*, 50(3-4):409–425, 2004.
15. M. Eslaminia and M. N. Guddati. A double-sweeping preconditioner for the Helmholtz equation. *J. Comput. Phys.*, 314:800–823, 2016.
16. M. J. Gander, I. G. Graham, and E. A. Spence. Applying GMRES to the Helmholtz equation with shifted Laplacian preconditioning: what is the largest shift for which wavenumber-independent convergence is guaranteed? *Numer. Math.*, 131(3):567–614, 2015.
17. M. J. Gander and F. Nataf. AILU: a preconditioner based on the analytic factorization of the elliptic operator. *Numer. Linear Algebra Appl.*, 7(7-8):505–526, 2000.
18. M. J. Gander and F. Nataf. An incomplete LU preconditioner for problems in acoustics. *J. Comput. Acoust.*, 13(3):455–476, 2005.
19. M. J. Gander and H. Zhang. A class of iterative solvers for the Helmholtz equation: factorizations, sweeping preconditioners, source transfer, single layer potentials, polarized traces, and optimized Schwarz methods. *SIAM Rev.*, 61(1):3–76, 2019.
20. I. G. Graham, E. A. Spence, and J. Zou. Domain decomposition with local impedance conditions for the Helmholtz equation with absorption. *SIAM J. Numer. Anal.*, 58(5):2515–2543, 2020.
21. T. Hagstrom and S. Kim. Complete radiation boundary conditions for the Helmholtz equation I: waveguides. *Numer. Math.*, 141(4):917–966, 2019.
22. F. Ihlenburg and I. Babuška. Dispersion analysis and error estimation of Galerkin finite element methods for the Helmholtz equation. *Internat. J. Numer. Methods Engrg.*, 38(22):3745–3774, 1995.
23. 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(9):9–37, 1995.
24. F. Ihlenburg and I. Babuška. Finite element solution of the Helmholtz equation with high wave number. II. The *h-p* version of the FEM. *SIAM J. Numer. Anal.*, 34(1):315–358, 1997.
25. S. Kim. Error analysis of PML-FEM approximations for the Helmholtz equation in waveguides. *ESAIM Math. Model. Numer. Anal.*, 53(4):1191–1222, 2019.
26. S. Kim. Analysis of complete radiation boundary conditions for the Helmholtz equation in perturbed waveguides. *J. Comput. Appl. Math.*, 367:112458, 2020.
27. S. Kim. Fractional order Sobolev spaces for the Neumann Laplacian and the vector Laplacian. *J. Korean Math. Soc.*, 57(3):721–745, 2020.
28. S. Kim and J. E. Pasciak. Analysis of a Cartesian PML approximation to acoustic scattering problems in $\mathbb{R}^2$. *J. Math. Anal. Appl.*, 370(1):168 – 186, 2010.
29. S. Kim and H. Zhang. Optimized double sweep Schwarz method with complete radiation boundary conditions for the Helmholtz equation in waveguides. *Comput. Math. Appl.*, 72(6):1573–1589, 2016.
30. W. Leng and L. Ju. An additive overlapping domain decomposition method for the Helmholtz equation. *SIAM J. Sci. Comput.*, 41(2):A1252–A1277, 2019.
31. J.-L. Lions and E. Magenes. *Non-homogeneous boundary value problems and applications. Vol. I.* Springer-Verlag, New York, 1972.

32. A. Modave, A. Royer, X. Antoine, and C. Geuzaine. A non-overlapping domain decomposition method with high-order transmission conditions and cross-point treatment for Helmholtz problems. *Comput. Methods Appl. Mech. Engrg.*, 368:113162, 2020.
33. F. Nataf and F. Nier. Convergence rate of some domain decomposition methods for overlapping and nonoverlapping subdomains. *Numer. Math.*, 75(3):357–377, 1997.
34. A. Schädle and L. Zschiedrich. Additive Schwarz method for scattering problems using the PML method at interfaces. In *Domain Decomposition Methods in Science and Engineering XVI*, volume 55 of *Lect. Notes Comput. Sci. Eng.*, pages 205–212. Springer, Berlin, 2007.
35. C. C. Stolk. A rapidly converging domain decomposition method for the Helmholtz equation. *J. Comput. Phys.*, 241:240 – 252, 2013.
36. C. C. Stolk. An improved sweeping domain decomposition preconditioner for the Helmholtz equation. *Adv. Comput. Math.*, 43(1):45–76, 2017.
37. C. C. Stolk, M. Ahmed, and S. K. Bhowmik. A multigrid method for the Helmholtz equation with optimized coarse grid corrections. *SIAM J. Sci. Comput.*, 36(6):A2819–A2841, 2014.
38. A. Toselli. Overlapping methods with perfectly matched layers for the solutions of the helmholtz equation. In *Eleventh International Conference on Domain Decomposition Methods,* C.H. Lai, P.E. Bjøstad, M. Cross and O.B. Widlund, eds., pages 551–558. DDM.org, 1990.
39. M. B. van Gijzen, Y. A. Erlangga, and C. Vuik. Spectral analysis of the discrete Helmholtz operator preconditioned with a shifted Laplacian. *SIAM J. Sci. Comput.*, 29(5):1942–1958, 2007.
40. P. Vaněk, J. Mandel, and M. Brezina. Two-level algebraic multigrid for the Helmholtz problem. In *Domain decomposition methods, 10 (Boulder, CO, 1997)*, volume 218 of *Contemp. Math.*, pages 349–356. Amer. Math. Soc., Providence, RI, 1998.
41. A. Vion and C. Geuzaine. Double sweep preconditioner for optimized Schwarz methods applied to the Helmholtz problem. *J. Comput. Phys.*, 266:171–190, 2014.
42. Zhiguo Yang, Li-Lian Wang, and Yang Gao. A Truly Exact Perfect Absorbing Layer for Time-Harmonic Acoustic Wave Scattering Problems. *SIAM J. Sci. Comput.*, 43(2):A1027–A1061, 2021.