

Convergence analysis of time-domain PMLs for 2D electromagnetic wave propagation in dispersive waveguides

Eliane Bécache, Maryna Kachanovska, Markus Wess

▶ **To cite this version:**

Eliane Bécache, Maryna Kachanovska, Markus Wess. Convergence analysis of time-domain PMLs for 2D electromagnetic wave propagation in dispersive waveguides. ESAIM: Mathematical Modelling and Numerical Analysis, 2023. hal-03910611

HAL Id: hal-03910611
https://hal.science/hal-03910611

Submitted on 22 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Will be set by the publisher

CONVERGENCE ANALYSIS OF TIME-DOMAIN PMLS FOR 2D ELECTROMAGNETIC WAVE PROPAGATION IN DISPERSIVE WAVEGUIDES

Eliane Becache[1], Maryna Kachanovska[1] and Markus Wess[2]

Abstract. This work is dedicated to the analysis of generalized perfectly matched layers (PMLs) for 2D electromagnetic wave propagation in dispersive waveguides. Under quite general assumptions on frequency-dependent dielectric permittivity and magnetic permeability we prove convergence estimates in homogeneous waveguides and show that the PML error decreases exponentially with respect to the absorption parameter and the length of the absorbing layer. The optimality of this error estimate is studied both numerically and analytically. Finally, we demonstrate that in the case when the waveguide contains a heterogeneity supported away from the absorbing layer, instabilities may occur, even in the case of the non-dispersive media. Our findings are illustrated by numerical experiments.

2020 Mathematics Subject Classification. 35L05, 35Q61, 65M12.

December 22, 2022.

Introduction

In this paper we are interested in 2D time-domain dispersive wave propagation in unbounded (or semi-bounded) waveguides.

For simulating wave propagation in unbounded domains using classical numerical approximation methods (finite differences, finite elements, Discontinuous Galerkin etc), the computational domain has to be artificially bounded. The Perfectly Matched Layer (PML) method, first introduced by Bérenger for Maxwell's equations (1994 [Ber94] in 2D, 1996 [Ber96] in 3D), is one of the most popular methods because of its efficiency and ease of implementation. The idea is to surround the computational domain with an absorbing layer (the PML region) so that the coupled system possesses the property of generating no reflections at the interface between the free medium and the artificial absorbing medium. PMLs have been widely used for the simulation of time dependent electromagnetic waves, as well as Helmholtz-like equations e.g., [Ged96, PZC98, TC98, ZC96, Ber97, TY98, Pet00]. The method has been then extended to various non-dispersive models (the paraxial wave equations [Col97], the linearized Euler equations [TAC98, Hu96, Hes98, DJ06a], elastic wave propagation in isotropic and anisotropic media [HSB96, CT01, BFJ03], etc.) and dispersive models [Cum04, BJK17, BJV18, BK17].

However, PMLs are known to produce instabilities when applied to some anisotropic [Hu96, BFJ03, Hu01, DI13, Nat06, AGH99] or dispersive [Cum04, BJV18, BK17, BJK17] media. Some of these have been overcome in the above-mentioned works, however, the question of stabilizing the PMLs remains model-dependent. Moreover, even in situations when the PMLs remain stable, their error control is rather difficult, because

Keywords and phrases: perfectly matched layers, time-dependent Maxwell equations, Laplace transform, waveguides, dispersive media, metamaterials

[1] POEMS, CNRS, Inria, ENSTA Paris, Institut Polytechnique de Paris, 828 Bd. des Maréchaux, 91120, Palaiseau, France

[2] Institute of Analysis and Scientific Computing, TU Wien, Vienna, Austria

© EDP Sciences, SMAI 1999

of the interplay of the various parameters of the PML and the discretization errors, see [CM98, MDG14, ADGK03, Che19].

Concerning the mathematical analysis of PMLs in the time-domain, there are mainly three questions of interest : (i) the well posedness, (ii) the stability and (iii) the convergence. It is important to precisely specify these notions: The PML model is well-posed if there exists a unique solution such that the L^2-norm of the solution can be bounded by some norm of the initial conditions multiplied by a constant $C(t)$ (which can possibly grow exponentially in time). The stronger notion of stability imposes that the constant $C(t)$ does not grow faster than some polynomial ($C(t) \leq Ct^k$). Finally, the question of convergence consists of estimating the error between the exact solution and the PML solution with respect to the PML parameters (typically parameters of the damping function and the length of the layer).

The question of the well-posedness of PMLs has been addressed very early in [AG97] for Maxwell's equations and has been then shown in various papers for several models including general hyperbolic systems and for a general non-constant damping parameters [AG97, AGH02, BJ02, BFJ03, AHK06, HPBR11, HJL18].

Concerning the stability, the situation is quite different and depends on the model considered. As previously mentioned, standard PMLs can give rise to instabilities for some anisotropic and/or dispersive models, as soon as there exist *backward waves* in the direction of the absorbing layer (i.e. waves for which the projections on the PML transverse direction of the phase and group velocities are of opposite signs [BFJ03]). In most of the papers the stability analysis is done in a simplified setting where all the absorption parameters are constant, based on Fourier or Laplace analysis or on energy techniques: see e.g. [BJ02, BPG04, BGIK19, AHK06, DI13] for the non-dispersive scalar wave equation or Maxwell's equation, [BFJ03, DK12] for elastic wave, [DJ06a, Hu01, Nat06] for aeroacoustics, and [BJK17, BJV18, BK17] for some dispersive models. Note that in [BFJ03, AHK06] general hyperbolic systems are also considered.

Up to our knowledge the only two works where the stability and convergence of PMLs for arbitrary (non-negative) absorption parameters were proven are the article by J. Diaz and P. Joly [DJ06b] and in our previous work [BK21]. In [DJ06b], the authors construct an explicit representation of the solution using the fundamental solution for the PML system for the 2D acoustic wave equation based on the Cagniard-de-Hoop contour deformation method and the method of reflections. Then they derive convergence estimates for the PMLs, which, according to the numerical experiments, are close to optimal. In [BK21], the case of a 3D acoustic waveguide is considered and again some convergence estimates are obtained from explicit representations of the solution, based on modal decompositions and some energy-like Laplace domain arguments.

In the present paper, we aim at extending the stability and convergence results obtained for a non-constant absorption parameter in [BK21] to the case of isotropic dispersive semi-infinite waveguides. A very important example of such media are metamaterials, *i.e.* artificial composite materials having extraordinary electromagnetic properties. In particular, Negative Index Metamaterials (NIMs), also called left-handed media or double negative metamaterials, have negative permittivity and permeability at some frequencies due to microscopic resonating structures [SPW04]. Since the 1990s, NIMs are the subject of active research due to their promising applications [CSL10]: superlensing, cloaking, improved antennae, etc. Even in isotropic media in the free space, NIMs naturally support backward waves (at least in some range of frequencies), which leads to instabilities of the standard PMLs. Some generalized PMLs have been introduced to overcome these instabilities, see [BJV18]. The key point is to introduce a damping parameter which depends on the frequency ω so that it can be adapted to the presence of backward waves. The stability of these generalized PMLs have been analysed for a constant absorption parameter (for the free space) in [BJV18] using Fourier techniques and in [BK17] using Laplace techniques.

We proceed as follows. Section 1 is dedicated to the presentation of the problem, and listing as well as deriving new properties of the dispersive media crucial for our analysis. We recall the generalized perfectly matched layers, introduced in [Cum04] and further generalized in [BJV18] and studied in [BK17]. Section 2 is of technical nature and is dedicated to a discussion of the well-posedness of the PML problem for the dispersive media, as well as deriving an explicit representation of the solution. Section 3 presents the convergence analysis. Section 4 is dedicated to the case when the waveguide is no longer homogeneous. We study the cases of waveguides composed of piecewise-constant non-dispersive and dispersive media, and, using semi-analytic arguments, demonstrate that in general, perfectly matched layers, even while being stable for the homogeneous media, may exhibit instabilities in the presence of the compact perturbations away from

the PMLs. Finally, we finish our exposition with numerical experiments in Section 5 and conclusions in Section 6.

1. Problem Setting

1.1. Physical preliminaries and the scope of the problem

We consider electromagnetic wave propagation in the two-dimensional semi-unbounded dispersive waveguide $\Omega = \mathbb{R}^+_* \times (0, \ell)$, $\ell > 0$, whose boundary is denoted by Γ and the respective exterior normal by $\boldsymbol{n} = (n_x, n_y)^T$. In particular, we study the TM system of Maxwell's equations for the electromagnetic field $\boldsymbol{E} = (E_x, E_y)^T : \mathbb{R}^+ \times \Omega \to \mathbb{R}^2$ and $H : \mathbb{R}^+ \times \Omega \to \mathbb{R}$:

$$\begin{aligned} &\partial_t \boldsymbol{D} - \mathbf{curl}\, H = 0, \\ &\partial_t B + \operatorname{curl} \boldsymbol{E} = 0, \quad \text{in } \mathbb{R}^+_* \times \Omega, \end{aligned} \tag{M}$$

which we later on equip with initial and boundary conditions (the results of the paper will extend naturally to the case when there are source terms in the formulation (M)).

The relations between the electric induction $\boldsymbol{D}$ and the electric field $\boldsymbol{E}$, and between the magnetic induction B and the magnetic field H are given in the Laplace domain. To describe them we introduce the Laplace transform, defined for causal (i.e., vanishing on $\mathbb{R}^-_*$) functions $\phi \in L^1(\mathbb{R})$ by

$$\hat{\phi}(s) = (\mathcal{L}\phi)(s) = \int_0^{+\infty} \mathrm{e}^{-st} \phi(t) dt, \quad s \in \mathbb{C}^+_* = \{z \in \mathbb{C} : \operatorname{Re} z > 0\}.$$

Provided the respective dielectric permittivity and magnetic permeability of the medium ε, $\mu : \mathbb{C}^+_* \to \mathbb{C}$, we equip the system (M) with the constitutive laws

$$\hat{\boldsymbol{D}}(s, \boldsymbol{x}) = \varepsilon(s) \hat{\boldsymbol{E}}(s, \boldsymbol{x}), \quad \hat{B}(s, \boldsymbol{x}) = \mu(s) \hat{H}(s, \boldsymbol{x}). \tag{1}$$

Rewriting the system (M) in the Laplace domain in its second order formulation for the field $\hat{H}$ yields in particular the scalar wave equation (where we assumed that the initial conditions vanish):

$$s^2 \mu(s) \varepsilon(s) \hat{H} - \Delta \hat{H} = 0 \text{ in } \Omega. \tag{2}$$

The dielectric permittivity and the magnetic permeability are such that in the time domain the constitutive relations (1) can be written in the form of convolutions of causal tempered distributions, which a priori requires that ε and μ are analytic in $\mathbb{C}^+_*$:

$$\boldsymbol{D} = \varepsilon(\partial_t) \boldsymbol{E} = \mathcal{L}^{-1}\big(\varepsilon(s) \hat{\boldsymbol{E}}\big), \quad B = \mu(\partial_t) H = \mathcal{L}^{-1}\big(\mu(s) \hat{H}\big). \tag{CR}$$

In the following paragraph we will discuss various properties that may be satisfied by ε and μ, which will be important at different stages of the analysis. Some of these properties have been already studied in the literature and/or have some physical meaning behind, cf. [CJK17, BK17], while some of them seem to be of a more technical nature.

Let us remark that here we work in the dimensionless setting of the Maxwell's equations, where the dielectric permittivity of the vacuum and its magnetic permeability are $\varepsilon_0 = \mu_0 = 1$.

1.1.1. *Properties of the dielectric permittivity, the magnetic permeability and the relation between them*

Let $\eta : \mathbb{C}^+_* \to \mathbb{C}$ denote ε or μ. We call a function η which is analytic in $\mathbb{C}^+_*$

- *time-real*, if $\overline{\eta(s)} = \eta(\overline{s})$, for all $s \in \mathbb{C}^+_*$.
  This property is related to the fact that $\mathcal{L}^{-1}\eta(s)$ is a *real-valued* causal distribution.
- *passive* if

$$\operatorname{Re}(s\eta(s)) > 0, \quad \text{for all} \quad s \in \mathbb{C}^+_*. \tag{P}$$

Passivity is related to the stability of the problem, as discussed in e.g. [CJK17, BK17]. When written with the Fourier transform convention, rather than the Laplace transform one, with $s = -i\omega$, it reduces to requiring that $\mathrm{Im}(\omega\eta(-i\omega)) > 0$ for $\omega \in \mathbb{C}$ with $\mathrm{Im}(\omega) > 0$, i.e. that $\omega\eta(-i\omega)$ is a Herglotz function.

- *non-dispersive at high frequencies* if, for all $\alpha \in (-\frac{\pi}{2}, \frac{\pi}{2})$, $\lim\limits_{R\to+\infty} \eta(R\mathrm{e}^{i\alpha}) = 1$.

The above properties have already appeared in the literature, cf. e.g., [CJK17] (with a slightly different formulation of the non-dispersivity at high frequencies).

Definition 1.1. We will call the function $\eta : \mathbb{C}_*^+ \to \mathbb{C}$ admissible if η is analytic in $\mathbb{C}_*^+$, time-real, passive and non-dispersive at high frequencies.

In the present paper we will limit our discussion to admissible materials.

Assumption 1.2. *The functions $s \mapsto \varepsilon(s)$ and $s \mapsto \mu(s)$ are admissible.*

It can be shown, see [CJK17], that this class of materials includes the media considered in the PhD thesis [Vin16], some results of which are summarized in the article [BJV18], where the perfectly matched layers were constructed for non-dissipative media. The idea of [BJV18] was extended to more general media, including in particular materials satisfying the above assumption, in [Vin16].

A standard example of such media is given by so-called Lorentz materials :

Definition 1.3. The medium (ε, μ) is called Lorentz, if

$$\varepsilon(s) = 1 + \sum_{n=0}^{N_\varepsilon} \frac{\varepsilon_j}{\omega_{\varepsilon,j}^2 + s^2}, \quad \varepsilon_j > 0, \quad \omega_{\varepsilon,j} \geq 0, \quad j = 0, \dots, N_\varepsilon, \quad N_\varepsilon \in \mathbb{N},$$
$$\mu(s) = 1 + \sum_{n=0}^{N_\mu} \frac{\mu_j}{\omega_{\mu,j}^2 + s^2}, \quad \mu_j > 0, \quad \omega_{\mu,j} \geq 0, \quad j = 0, \dots, N_\mu, \quad N_\mu \in \mathbb{N}.$$

In the particular case when $N_\varepsilon = N_\mu = 0$ and $\omega_{\mu,0} = \omega_{\varepsilon,0} = 0$, the corresponding medium is called Drude.

Lorentz media are non-dissipative: plane waves propagate rather than decay in such media.

Example 1.4. Other examples of dispersive media are given e.g. by

- Havriliak-Negami dielectrics (see [HN67]):

$$\varepsilon(s) = 1 + \frac{\varepsilon_0}{\left(1 + (s\tau)^\alpha\right)^\beta}$$

  with $\tau, \alpha, \beta > 0$, which reduce to Cole-Cole dielectrics (see [CC41]) for $\beta = 1$, and
- dissipative Drude materials:

$$\varepsilon(s) = 1 + \frac{\varepsilon_0}{s^2 + \gamma_{\varepsilon,0} s}, \quad \mu(s) = 1 + \frac{\mu_0}{s^2 + \gamma_{\mu,0} s}, \quad \gamma_{\varepsilon,0}, \gamma_{\mu,0} > 0.$$

It can be shown that any (even dissipative) medium that satisfies Assumption 1.2 can be viewed as a generalization of the Lorentz materials. More precisely, repeating the proof of Theorem 4.5 in [CJK17] almost verbatim, one obtains the following refinement of the famous Herglotz-Nevanlinna representation theorem.

Theorem 1.5 (cf. Theorem 4.5 in [CJK17])**.** *Let $\eta : \mathbb{C}_*^+ \to \mathbb{C}$ be analytic, passive, time-real and satisfy $\lim\limits_{r\to+\infty} \eta(r) = 1$. Then there exists a Borel regular measure ν s.t.*

$$\eta(s) = 1 + \int_{\mathbb{R}} \frac{d\nu(\xi)}{\xi^2 + s^2}, \quad s \in \mathbb{C}_*^+, \qquad \textit{with } \int_{\mathbb{R}} d\nu(\xi)/(1+\xi^2) < \infty.$$

An immediate corollary of the above is

Lemma 1.6. *Let analytic $\eta : \mathbb{C}_*^+ \to \mathbb{C}$ be admissible (cf. Defn. 1.1). Then*

$$\mathrm{Re}(s\eta(s)) \geq \mathrm{Re}s, \quad s \in \mathbb{C}_*^+. \tag{3}$$

Another technical property that we encountered during the analysis is stated in the lemma below.

Lemma 1.7. *Let analytic $\eta : \mathbb{C}_*^+ \to \mathbb{C}$ be admissible (cf. Defn. 1.1). Then*

$$\mathrm{Re}(s\eta(s)) \geq \mathrm{Re}(s) \max\left\{|\eta(s)|, 2 - |\eta(s)|\right\}, \quad \textit{for all } s \in \mathbb{C}_*^+. \tag{MP}$$

Proof. First of all, by Theorem 1.5,

$$\mathrm{Re}(s\eta(s)) = \mathrm{Re}s \left(1 + \int_{\mathbb{R}} \frac{|s|^2 + \xi^2}{|s^2 + \xi^2|^2} d\nu(\xi)\right). \tag{4}$$

A straightforward computation yields

$$\left|\int_{\mathbb{R}} \frac{d\nu(\xi)}{s^2 + \xi^2}\right| = \left|\int_{\mathbb{R}} \frac{\bar{s}^2 + \xi^2}{|s^2 + \xi^2|^2} d\nu(\xi)\right| \leq \int_{\mathbb{R}} \frac{|s|^2 + \xi^2}{|s^2 + \xi^2|^2} d\nu(\xi) =: p(s).$$

Obviously,

$$1 - p(s) \leq |\eta(s)| \leq 1 + p(s), \quad \text{and also} \quad 2 - |\eta(s)| \leq 1 + p(s).$$

The above inequalities combined with (4) yield the desired statement of the lemma. □

1.1.2. *Stability of the problem (M, CR)*

First of all, let us rewrite the problem (M, CR) in a more convenient form. We introduce the polarization $\boldsymbol{P}$ and the magnetization M, related to $\boldsymbol{E}$, H via the identities

$$\hat{\boldsymbol{P}} = \chi_e(s)\hat{\boldsymbol{E}}, \quad \hat{M}_z = \chi_m(s)\hat{H}, \quad \chi_e(s) = \varepsilon(s) - 1, \quad \chi_m(s) = \mu(s) - 1. \tag{5}$$

This allows to rewrite the problem (M), (CR), which we equip with initial and boundary conditions, in the following form:

$$\begin{aligned}
&\partial_t \boldsymbol{E} + \partial_t \boldsymbol{P} - \mathbf{curl}\, H = 0,\\
&\partial_t H + \partial_t M + \operatorname{curl} \boldsymbol{E} = 0, \quad \text{in } \mathbb{R}_*^+ \times \Omega,\\
&E_x n_y - E_y n_x = 0, \quad \text{on } \mathbb{R}_*^+ \times \Gamma,\\
&\boldsymbol{E}(0, \boldsymbol{x}) = \boldsymbol{E}_0(\boldsymbol{x}), \quad H(0, \boldsymbol{x}) = H_0(\boldsymbol{x}), \quad \boldsymbol{P}(0, \boldsymbol{x}) = 0, \quad M(0, \boldsymbol{x}) = 0 \quad \text{in } \Omega.
\end{aligned} \tag{6}$$

We recall the following result, which is a reformulation of [CJK17, Theorem 4.8] and its proof, see also [BK21].

Proposition 1.8 ([CJK17])**.** *Under Assumption 1.2, for every $(\boldsymbol{E}_0, H_0) \in \left(L^2(\Omega)\right)^3$, for all $T > 0$, the problem (6), (5) has a unique solution $(\boldsymbol{E}, H) \in L^2(0, T; \left(L^2(\Omega)\right)^3)$ and $(\boldsymbol{P}, M)$ in a class causal tempered $(L^2(\Omega))^3$-valued distributions. Moreover, for any sufficiently regular solution, the electromagnetic energy is uniformly bounded:*

$$\mathcal{E}(t) = \frac{1}{2}\left(\|\boldsymbol{E}(t)\|^2 + \|H(t)\|^2\right) \leq \mathcal{E}(0), \quad t \in (0, T).$$

The difficulty in the numerical simulation of (5,6) lies in the unboundedness of the domain. To deal with it, we will use the perfectly matched layer method.

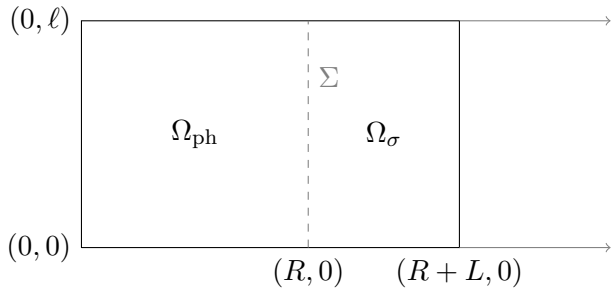


FIGURE 1. Sketch of the decomposition of the waveguide into the physical domain Ω_{ph}, PML domain Ω_σ, and the computational domain $\Omega_c = \text{int}(\overline{\Omega_{ph}} \cup \overline{\Omega_\sigma})$.

1.2. **Perfectly Matched Layer Method**

1.2.1. *Generalized PMLs*

It was demonstrated in [Cum04] that Bérenger's classical perfectly matched layers can exhibit instabilities when applied to dispersive media. Such instabilities are induced by the presence of backward propagating waves, as explained in [BJV18]. A remedy, originally suggested in [Cum04] for Drude materials, and further generalized in [BJV18, BK17], to treat the cases of more general media is altering the perfectly matched layer change of variables in order to account for the dispersion of the problem. Let us describe the respective procedure in more detail.

First of all, we assume that

Assumption 1.9. $\operatorname{supp} \boldsymbol{E}_0, \operatorname{supp} H_0 \subset [0,R) \times (0,\ell)$.

Our goal is to compute the solution to the original problem (6) in the physical domain $\Omega_{ph} := (0,R)\times(0,\ell)$, by constructing a perfectly matched layer at $x = R$ of width L. For this we first rewrite the problem (6), (5) in the Laplace domain, which yields, after eliminating $\boldsymbol{P}$ and M_z,

$$\begin{aligned} &s\varepsilon(s)\hat{\boldsymbol{E}} - \boldsymbol{E}_0 - \mathbf{curl}\,\hat{H} = 0,\\ &s\mu(s)\hat{H} - H_0 + \operatorname{curl}\hat{\boldsymbol{E}} = 0, \qquad \text{in } \Omega,\\ &\hat{E}_x n_y - \hat{E}_y n_x = 0 \text{ on } \Gamma. \end{aligned}$$

Next, let us define the absorption function $\sigma : (0, R+L) \to \mathbb{R}^+$ satisfying

$$\sigma \in L^\infty(0,R+L) \text{ and } \sigma = 0 \text{ on } (0,R). \tag{7}$$

With this definition, the [BJV18]-PML change of variables reads

$$\widetilde{x} = x + \frac{1}{s\zeta(s)} \int_R^x \sigma(x')dx', \tag{8}$$

where the function ζ is analytic in $\mathbb{C}_*^+$, and is responsible for stabilizing the PMLs for dispersive media. Note that in the classical Bérenger PML $\zeta(s) = 1$. We also introduce the average damping $\bar{\sigma}$ by

$$\bar{\sigma} = \frac{1}{L} \int_R^{L+R} \sigma(x)dx. \tag{9}$$

The interface between the PML and the physical media will be denoted by $\Sigma = \{R\} \times (0,\ell)$, and the PML media by $\Omega_\sigma = (R, R+L) \times (0,\ell)$. See Figure 1 for an illustration of the geometry.

Remark 1.10. In [BJV18, BK17], the PML change of variables (8) was written in the form

$$\widetilde{x} = x + \frac{\psi(s)}{s}\int_R^x \sigma(x')dx',$$

i.e., with $\psi = \zeta^{-1}$. Because we will limit our discussion to the choices of $\psi(s)$ suggested in [BJV18, BK17], we will take ψ, s.t. ψ^{-1} is admissible in the sense of Definition 1.1. Thus we feel that it is more natural to introduce $\zeta = \psi^{-1}$.

After performing this change of variables and truncating the PML system, we look for $(\hat{\boldsymbol{E}}^\sigma, \hat{H}^\sigma)$ that solve the following problem in the computational domain $\Omega_c = (0, L+R)\times(0,\ell)$, with the boundary Γ_c:

$$\begin{aligned}
&s\varepsilon(s)\hat{E}_x^\sigma - \partial_y \hat{H}^\sigma = E_{0x}, \quad \text{in } \Omega_c,\\
&s\varepsilon(s)\left(1+\frac{\sigma}{s\zeta(s)}\right)\hat{E}_y^\sigma + \partial_x \hat{H}^\sigma = E_{0y}, \quad \text{in } \Omega_c,\\
&s\mu(s)\hat{H}^\sigma + \left(1+\frac{\sigma}{s\zeta(s)}\right)^{-1}\partial_x \hat{E}_y^\sigma - \partial_y \hat{E}_x^\sigma = H_{z0}, \quad \text{in } \Omega_c,\\
&\hat{E}_x^\sigma n_y - \hat{E}_y^\sigma n_x = 0 \text{ on } \Gamma_c.
\end{aligned} \tag{10}$$

It remains to rewrite the above problem in the time domain. For this, however, we need to choose a function ζ that ensures that the time-domain counterpart of the above system is stable.

1.2.2. *Choice of ζ for any media satisfying Assumption 1.2 and the corresponding PML system in the time-domain*

In [BJV18, BK17] a choice of ζ, depending on ε, μ, was proposed based on the stability analysis performed for the case when $\sigma = \text{const} > 0$ in $(0, R+L)$ (or even in the case when $\Omega_c = \mathbb{R}^d$, cf. [BJV18]). Numerical experiments indicate that the time-domain counterpart of (10), provided the choice of ζ based on the analysis for $\sigma = \text{const} > 0$, is stable as well. Therefore, we will limit our choice of the parameter ζ to the one suggested in [BJV18, BK17], more precisely, to the one described in [BK17, Theorem 3.7], well adapted to our setting.

Assumption 1.11. *ζ is an analytic in $\mathbb{C}_*^+$ function which is admissible in the sense of Defn. 1.1 and*

$$\mathrm{Re}(s\mu(s)\varepsilon(s)\zeta(s)^{-1}) > 0, \quad s \in \mathbb{C}_*^+,$$

i.e., $\mu(s)\varepsilon(s)\zeta(s)^{-1}$ is also admissible.

Evidently, $\zeta = \mu$ or $\zeta = \varepsilon$ satisfy the above assumption. The advantage in choosing different functions ζ lies in the possible reduction of the number of auxiliary unknowns required for the implementation of the PML system. Let us remark that, as shown in [BK17], this choice of ζ covers the stabilizing functions suggested in [BJV18] for Lorentz materials.

We rewrite the PML system (10) in the time-domain, for $(t, \boldsymbol{x}) \in \mathbb{R}_*^+ \times \Omega_c$,

$$\begin{aligned}
&\partial_t E_x^\sigma + \partial_t P_x^\sigma - \partial_y H^\sigma = 0,\\
&\partial_t E_y^\sigma + \partial_t P_y^\sigma + \partial_x H^\sigma - \sigma\widetilde{E}_y^\sigma = 0,\\
&\partial_t \widetilde{E}_y^\sigma + \partial_t \widetilde{P}_y^\sigma + \sigma\widetilde{E}_y^\sigma - \partial_x H^\sigma = 0,\\
&\partial_t H^\sigma + \partial_t M^\sigma - \sigma\widetilde{H}_y^\sigma + \partial_x E_y^\sigma - \partial_y E_x^\sigma = 0,\\
&\partial_t \widetilde{H}_y^\sigma + \partial_t \widetilde{M}_y^\sigma + \sigma\widetilde{H}_y^\sigma = \partial_x E_y^\sigma,
\end{aligned} \tag{11}$$

equipped with the constitutive relations in $\mathbb{R}_*^+ \times \Omega_c$:

$$\boldsymbol{P}^\sigma = \chi_e(\partial_t)\boldsymbol{E}^\sigma, \quad M^\sigma = \chi_m(\partial_t)H^\sigma, \quad \widetilde{M}_y^\sigma = \chi_\zeta(\partial_t)\widetilde{H}_y^\sigma, \quad \widetilde{P}_y^\sigma = \chi_\zeta(\partial_t)\widetilde{E}_y^\sigma, \tag{12}$$

with $\chi_\zeta(s) = \zeta(s) - 1$, initial conditions in Ω_c:

$$\boldsymbol{E}^\sigma(0) = \boldsymbol{E}_0, \quad H_z^\sigma(0) = H_{z0}, \quad \boldsymbol{P}^\sigma(0) = 0, \quad M^\sigma(0) = 0, \quad \widetilde{H}_y^\sigma(0) = \widetilde{M}_y^\sigma(0) = 0, \tag{13}$$

and the boundary conditions

$$E_x^\sigma n_y - E_y^\sigma n_x = 0 \text{ on } \mathbb{R}_*^+ \times \Gamma_c. \tag{14}$$

For the well-posedness and stability analysis, we will not work with (11,12,13,14) directly, but rather with its Laplace domain counterpart. In particular, we start with (10), and express all the unknowns with the help of $\hat{H}^\sigma$, which yields the following 2D Helmholtz equation with the PML change of variables:

$$\begin{aligned} &s^2\mu\varepsilon\left(1+\frac{\sigma}{s\zeta}\right)\hat{H}^\sigma - \partial_x\left(1+\frac{\sigma}{s\zeta}\right)^{-1}\partial_x\hat{H}^\sigma - \left(1+\frac{\sigma}{s\zeta}\right)\partial_y^2\hat{H}^\sigma = \hat{f} \quad \text{in } \Omega_c, \\ &\nabla\hat{H}^\sigma \cdot \boldsymbol{n} = 0 \quad \text{on } \Gamma_c, \end{aligned} \tag{15}$$

where $\hat{f} = s\varepsilon(s)H_{z0} - \operatorname{curl}\boldsymbol{E}_0$.

1.3. **Principal result of the paper**

The principal goal of this paper is to prove convergence of the perfectly matched layers in homogeneous waveguides and the main result Theorem 1.12 is given below. We formulate it with respect to the field H; for the field $\boldsymbol{E}$ the respective results can be obtained similarly, but for the price of some additional technical computations.

Theorem 1.12. *Let $T > 0$, H be a sufficiently regular solution of (6) with the initial data satisfying Assumption 1.9, for $0 < t < T$, more precisely, $H \in W^{1,2}(0,T;H^1(\Omega))$.*

Let μ, ε be admissible materials in the sense of Definition 1.1 and ζ be chosen such that Assumption 1.11 holds. Let $L > 0$ and σ satisfy (7). Then a unique solution of the PML problem (11-14) $H^\sigma \in L^\infty(0,T;L^2(\Omega_{ph}))$ satisfies the following error bound, with $e^\sigma = H^\sigma - H$:

$$\|e^\sigma\|_{L^\infty(0,T;L^2(\Omega_{ph}))} \begin{cases} = 0, & T \in [0, 2L), \\ \leq C\mathrm{e}^{-4L^2\bar{\sigma}T^{-1}}\frac{T^{3/2}}{\bar{\sigma}L^2}\max(1, \frac{T}{\bar{\sigma}L^2})\|\gamma_0\partial_t H\|_{L^2(0,T;L^2(\Sigma))}, & T \geq 2L, \end{cases}$$

where the constant $C > 0$ in the inequality is independent of $T, L, \bar{\sigma}, \Omega_{ph}$ and the solution of (6).

This result is in accordance with or extends known results: in [DJ06b] the authors prove a similar exponential bound with respect to the PML parameter and length for Cartesian perfectly matched layers applied to the acoustic wave propagation in vacuum. Since Theorem 1.12 includes also the case of non-dispersive materials (i.e., $\mu = \varepsilon \equiv 1$) it extends the bound from [BK21] to the case of dispersive waveguides. Note that the factor 4 in the exponential in the work at hand refines the bound from [BK21] to the best achievable one.

Remarkably the exponential convergence of the PML truncation is independent of the specific materials μ, ε and the choice of the function ζ. However, the choice of ζ (cf. [BJV18]) greatly influences the ease of implementation of the PML system. On the other hand, as argued in the recent article [BDT21] for the case of the Klein-Gordon equation, different choices of ζ behave differently if long-time errors are considered; unfortunately, the above error bound does not capture this effects.

2. Well-posedness and explicit representation of the solution

We analyze the stability and the convergence of the PML based on the Laplace-domain techniques, as suggested e.g. in [BK21]. For this recall the class $TD(X)$ defined by F. Sayas [Say16].

Definition 2.1 (cf. [Say16])**.** Let X be a Banach space. Then the class $TD(X)$ consists of causal (i.e., vanishing on $(-\infty, 0)$) X-valued distributions, s.t. for each $\Phi \in TD(X)$, there exists a causal (i.e. vanishing on $(-\infty, 0]$) continuous function $\phi : \mathbb{R} \to X$ and constants C, p, $m \geq 0$, s.t. for all $t \geq 0$,

$$\|\phi(t)\| \leq C(1 + t^p), \qquad \text{and } \Phi = \frac{d^m \phi}{dt^m}.$$

Here $\frac{d}{dt}$ is understood as a distributional derivative.
Such distributions are characterized by bounds on their Laplace transforms.

Theorem 2.2 (Propositions 3.1.1, 3.1.2, 3.1.3 in [Say16])**.** *A function $\Phi : \mathbb{C}_*^+ \to X$ is a Laplace transform of $\phi \in TD(X)$ if and only if:*

(1) Φ is holomorphic in $\mathbb{C}_*^+$, and
(2) Φ satisfies the following bound in $\mathbb{C}_*^+$:

$$\|\Phi(s)\| \leq |s|^\mu C_\Phi(\mathrm{Re} s), \qquad \mu \in \mathbb{R}, \tag{16}$$

where $C_\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ is non-increasing and satisfies, for some $m \geq 0$ and $C > 0$,

$$C_\Phi(\eta) \leq C\eta^{-m}, \text{ for all } \eta \in (0, 1].$$

Therefore, we start by proving the well-posedness of the following problem: given $\hat{f} \in \widetilde{H}^{-1}(\Omega_c) = (H^1(\Omega_c))'$, find $\hat{H}^\sigma \in H^1(\Omega_c)$ that satisfies (15), for $s \in \mathbb{C}_*^+$. Moreover, per Theorem 2.2, the solution to this problem should satisfy the stability bound

$$\|\hat{H}^\sigma\|_{H^1(\Omega_c)} \lesssim |s|^m \min(1, \mathrm{Re} s)^{-\nu} \|\hat{f}\|_{\widetilde{H}^{-1}(\Omega_c)}, \quad m \in \mathbb{Z},\ \nu \geq 0. \tag{17}$$

These results are achieved by proving an inf-sup condition satisfied by the sesquilinear form associated to (15) (let us remark that it relies on some non-local in time and space constructions, and thus cannot be easily applied to waveguides with perturbations). The bound (17) implies in particular that with $\hat{f}$ given after equation (15), the solution to (15) is a causal tempered distribution, and, in particular, a distributional derivative of some order of a continuous causal function of polynomial growth. The time-domain stability estimates on this function can be obtained by e.g. the Plancherel theorem or contour integration techniques, see [Say16] for more details. The drawback of this approach is that the obtained stability bounds are quite far from optimal. Significantly finer stability estimates on the solution H^σ can be obtained by exploiting an explicit representation of the solution and the Plancherel theorem, cf. [BK21]. This is the point of view we are going to pursue in this article.

2.1. **Existence and uniqueness for the PML problem** (15)

Let us consider the following problem: given $p \in \widetilde{H}^{-1}(\Omega_c)$, find $u \in H^1(\Omega_c)$ that satisfies

$$\begin{aligned} &a_s(u, v) = \langle p, v \rangle, \quad v \in H^1(\Omega_c), \\ &a_s(u, v) = s^2 \mu(s) \varepsilon(s) \int_{\Omega_c} \left(1 + \frac{\sigma}{s\zeta(s)}\right) u\, \overline{v} + \int_{\Omega_c} \left(1 + \frac{\sigma}{s\zeta(s)}\right)^{-1} \partial_x u\, \partial_x \overline{v} + \int_{\Omega_c} \left(1 + \frac{\sigma}{s\zeta(s)}\right) \partial_y u\, \partial_y \overline{v}. \end{aligned} \tag{18}$$

We have the following result, which generalizes Lemma 3.6 in [BK21].

Lemma 2.3. *For any $s \in \mathbb{C}_*^+$, the form $a_s : H^1(\Omega_c) \times H^1(\Omega_c) \to \mathbb{C}$ is continuous, and satisfies the following estimate:*

$$\inf_{u \in H^1(\Omega_c)} \sup_{v \in H^1(\Omega_c)} \frac{|a_s(u, v)|}{\|u\|_{H^1(\Omega_c)} \|v\|_{H^1(\Omega_c)}} \geq c_s,$$

where $c_s = c(\sigma) \frac{\min(1, (\mathrm{Re} s)^\ell)}{|s|^k}$, for some ℓ, $k \geq 0$.

Proof. See Appendix A. □

As discussed in the beginning of Section 2, this implies in particular that the respective PML system admits a unique solution. Let us denote by $\boldsymbol{U} = (E_x^\sigma,\, E_y^\sigma,\, \widetilde{E}_y^\sigma,\, H^\sigma,\, \widetilde{H}_y^\sigma)$, $\widetilde{\boldsymbol{U}} = (P_x^\sigma,\, P_y^\sigma,\, \widetilde{P}_y^\sigma,\, M^\sigma,\, \widetilde{M}_y^\sigma)$ the solution of (11).

Corollary 2.4. *The PML system* (11)*, equipped with the boundary conditions* (14)*, constitutive relations* (12)*, initial conditions* $\boldsymbol{U}|_{t=0} = \boldsymbol{U}_0 \in (L^2(\Omega_c))^5$ *and vanishing initial conditions for the rest of unknowns, admits a unique solution* $(\boldsymbol{U}, \widetilde{\boldsymbol{U}}) \in TD((H^1(\Omega_c))^{10})$.

Remark that the result holds true for initial conditions supported inside Ω_c. Moreover, one can show some a priori stability bounds, by proving that $\boldsymbol{U} = \frac{d^m}{dt^m}\boldsymbol{u}$ (where d/dt is a distributional derivative), with $\boldsymbol{u}$ being a continuous causal function, s.t. $\|\boldsymbol{u}(t)\|_{H^1(\Omega_c)} \lesssim (1+t)^k \|\boldsymbol{U}_0\|_{L^2(\Omega_c)}$, cf. the techniques of Chapter 3 of [Say16].

2.2. An explicit solution in the Laplace domain for the data supported in the physical domain

Our analysis relies on the modal decomposition. In particular, we denote by $(\lambda_m^2, \phi_m(y))$ the eigenvalues and eigenfunctions of the transverse Neumann Laplacian:

$$\begin{aligned} &-\partial_y^2\phi_m = \lambda_m^2\phi_m, \quad \partial_y\phi_m(0) = \partial_y\phi_m(\ell) = 0,\\ &0 < \lambda_0^2 \le \lambda_1^2 \le \ldots \to +\infty, \qquad \|\phi_m\|_{L^2(0,\ell)} = 1. \end{aligned}$$

For $u \in L^2(\Omega_c)$, let us denote by $u_m(x)$ its decomposition into Fourier series in ϕ_m: $u(x,y) = \sum\limits_{m=0}^{\infty} u_m(x)\phi_m(y)$.

2.2.1. *The symbol of the PML DtN map*

Using the same argument as in the proof of Lemma 2.3 it follows that the following problem: given $g \in H^{1/2}(\Sigma)$, find $u \in H^1(\Omega_\sigma)$ s.t.

$$s^2\mu\varepsilon\left(1+\frac{\sigma}{s\zeta}\right)u - \partial_x\left(1+\frac{\sigma}{s\zeta}\right)^{-1}\partial_x u - \left(1+\frac{\sigma}{s\zeta}\right)\partial_y^2 u = 0 \quad \text{in } \Omega_\sigma, \tag{19}$$

$$\nabla u\cdot\boldsymbol{n} = 0 \quad \text{on } \partial\Omega_\sigma\setminus\Sigma, \quad u = g, \quad \text{on } \Sigma, \tag{20}$$

is well-posed. We call the symbol of the PML DtN map the operator $T^\sigma : H^{1/2}(\Sigma) \mapsto (H^{1/2}(\Sigma))'$, defined for sufficiently regular functions u as follows:

$$T^\sigma g = \gamma_1 u(R,y),$$

where γ_1 is a conormal derivative on Σ, defined for sufficiently regular σ (e.g. piecewise-constant) and u by

$$\gamma_1 u(R,y) = \left(1+\frac{\sigma(L^+)}{s}\right)^{-1}\partial_x u(R,y).$$

To obtain an explicit representation of the DtN we apply the modal decomposition to (19), which yields the family of one-dimensional problems: find $u_m \in H^1(R, R+L)$, s.t.

$$\begin{aligned} &s^2\mu\varepsilon\left(1+\frac{\sigma}{s\zeta}\right)u_m - \partial_x\left(1+\frac{\sigma}{s\zeta}\right)^{-1}\partial_x u_m + \left(1+\frac{\sigma}{s\zeta}\right)\lambda_m^2 u_m = 0 \quad \text{in } (R, R+L),\\ &\partial_x u_m(R+L) = 0, \quad u_m(R) = g_m. \end{aligned} \tag{21}$$

A computation similar to [BK21] yields with

$$\kappa(s,\lambda) = \sqrt{s^2\varepsilon(s)\mu(s) + \lambda^2}, \tag{22}$$

the identity

$$\gamma_1 u_j(R) = T_j^\sigma(s) g_j, \quad T_j^\sigma(s) = -\kappa(s,\lambda_j)\frac{1-\exp(-2\kappa(s,\lambda_j)\gamma(s))}{1+\exp(-2\kappa(s,\lambda_j)\gamma(s))},$$

where $\gamma(s)$ is defined by

$$\gamma(s) = L + \frac{1}{s\zeta(s)}\int_R^{L+R}\sigma(x)dx = L\left(1+\frac{\bar{\sigma}}{s\zeta(s)}\right). \tag{23}$$

This allows us to define the symbol of the PML-DtN map by

$$T^\sigma(s)g = \sum_{m=0}^{\infty} T_m^\infty(s)\frac{1-\mathrm{e}^{-2\kappa(s,\lambda_m)\gamma(s)}}{1+\mathrm{e}^{-2\kappa(s,\lambda_m)\gamma(s)}} g_m \phi_m,$$

where $T_m^\infty(s) = -\kappa(s,\lambda_m)$ is the symbol of the modal radiating DtN map.

2.2.2. *An explicit expression for the solution*

Any solution of (15) satisfies the following problem (recall that supp $\hat{f} \subset (0,R)$):

$$\begin{aligned} &s^2\varepsilon\mu\hat{H}^\sigma - \Delta\hat{H}^\sigma = \hat{f}, \quad \text{in } \Omega_{ph},\\ &\nabla\hat{H}^\sigma\cdot\boldsymbol{n} = 0 \text{ on } \partial\Omega_{ph}\setminus\Sigma, \qquad \nabla\hat{H}^\sigma\cdot\boldsymbol{n} = T^\sigma\hat{H}^\sigma \text{ on } \Sigma. \end{aligned}$$

On the other hand, the exact solutions satisfies

$$\begin{aligned} &s^2\varepsilon\mu\hat{H} - \Delta\hat{H} = \hat{f} \quad \text{in } \Omega_{ph},\\ &\nabla\hat{H}\cdot\boldsymbol{n} = 0 \text{ on } \partial\Omega_{ph}\setminus\Sigma, \qquad \nabla\hat{H}\cdot\boldsymbol{n} = T^\infty\hat{H} \text{ on } \Sigma. \end{aligned}$$

Hence the error $\hat{e}^\sigma = \hat{H}^\sigma - \hat{H}$ satisfies

$$s^2\varepsilon\mu\hat{e}^\sigma - \Delta\hat{e}^\sigma = 0 \quad \text{in } \Omega_{ph}, \tag{24}$$

$$\nabla\hat{e}^\sigma\cdot\boldsymbol{n} = 0 \text{ on } \partial\Omega_{ph}\setminus\Sigma, \qquad \nabla\hat{e}^\sigma\cdot\boldsymbol{n} = T^\sigma\hat{e}^\sigma + (T^\sigma - T^\infty)\hat{H}. \tag{25}$$

The (unique) solution to the above problem can be found explicitly by using the modal decomposition (cf. [BK21] for a similar computation), which yields the explicit expression of the error

$$\hat{e}^\sigma = \sum_{m=0}^{\infty}\phi_m(y)\hat{e}_m^\sigma, \quad \hat{e}_m^\sigma(s,x) = c_m(s,x)\mathrm{e}^{-2\kappa(s,\lambda_m)\gamma(s)}, \tag{26}$$

$$c_m(s,x) = \left(1-\mathrm{e}^{-2\kappa(s,\lambda_m)(\gamma(s)+R)}\right)^{-1}\left(\mathrm{e}^{\kappa(s,\lambda_m)(x-R)} + \mathrm{e}^{-\kappa(s,\lambda_m)(x+R)}\right)\hat{H}_m(s,R).$$

We give the expression for the error between the exact solution and the PML solution in this way, since the term $\mathrm{e}^{-2\kappa(s,\lambda_m)\gamma}$ plays an important role in the convergence estimates.

Remark 2.5. Our goal is to obtain $L^\infty(0,T;L^2(\Omega))$ estimates on the error between the exact solution and the PML solution. If we wanted to obtain the $L^\infty(0,T;H^1(\Omega))$-norm of the error, it would have been simpler to rewrite the above expression by introducing auxiliary volume problems as it was done [BK21]. Since we want to keep the presentation as simple as possible, we stick to the above expression and concentrate on the estimate of the $L^\infty(0,T;L^2(\Omega))$-norms.

Let us remark that (26) indeed defines a Laplace transform of a TD($H^1(\Omega^c)$)-distribution of e^σ.

3. Convergence analysis

In this section we present the proof of the main result of this work Theorem 1.12. The proof of this theorem can be done in at least either of the two ways (cf. a similar discussion in [BK21]):

Bromwich formula and contour deformation: An application of the Bromwich formula for the inverse Laplace transform to the Laplace-domain error e^σ yields

$$e^\sigma(t,x,y) = \frac{1}{2\pi i}\sum_{m=0}^{\infty}\phi_m(y)\int_{\eta+i\mathbb{R}} \mathrm{e}^{st-2\kappa(s,\lambda_m)\gamma(s)}c_m(s,x)ds,$$

for every $\eta>0$. Recall that $\gamma(s)=L\big(1+\frac{\bar\sigma}{\zeta(s)}\big)$. Since we want to prove convergence for $\bar\sigma, L\to+\infty$, and c_m is bounded for large $\bar\sigma, L$, the only way to control the exponential growth of the integrand with respect to t is to bound the expression $st-2\kappa(s,\lambda_m)\gamma(s)$ on the integration contour.

The main idea is then to use the fact that the integrand is holomorphic and deform the integration contour from the straight vertical line $\eta+i\mathbb{R}$ to a contour $\mathcal{C}_m(\omega)=i\omega+\eta_m(\omega)$, $\omega\in\mathbb{R}$, for a smooth, bounded function $\eta_m>0$ in a way that the expression $\sup\limits_{s\in\mathcal{C}_m}|e^{st-2\gamma\kappa(s,\lambda_m)}|$ is minimized. Such a procedure would yield the bound

$$\begin{aligned}\left|\int_{\eta+i\mathbb{R}} \mathrm{e}^{st-2\kappa(s,\lambda_m)\gamma}c_m(s,x)ds\right| &= \left|\int_{\mathcal{C}_m} \mathrm{e}^{st-2\kappa(s,\lambda_m)\gamma}c_m(s,x)ds\right| \\ &\le \sup_{s\in\mathcal{C}_m}\left|e^{st-2\gamma\kappa(s,\lambda_m)}\right|\int_{\mathcal{C}_m}|c_m(s,x)||ds|. \end{aligned} \tag{27}$$

Assuming that the contour is regular enough, to bound the remaining integral, one can proceed similarly to the proof of [Say16, proof of Prop. 3.2.2], which exploits the regularity of the time-domain counterpart of c_m; this procedure would allow to obtain a bound on the error in terms of a certain $W^{k,1}$-norm of the time-domain c_m.

Plancherel's identity: Alternatively, we can use Plancherel's identity

$$\int_0^\infty \exp(-2\eta t)\,\|g(t)\,\|^2dt = \frac{1}{2\pi i}\int_{i\mathbb{R}+\eta}\|\hat g(s)\,\|^2ds.$$

for $g\in L^2(\mathbb{R}_+,X)$ on some Banach space X and $\eta>0$. Proceeding as in [BK21, proof of Lemma 4.3] enables us to obtain the bound of the following form, for some $\eta_m>0$,

$$\|e^\sigma\|_{L^2(0,T;L^2(\Omega))}\le C(T)\sup_{\lambda_m}\sup_{s\in\eta_m+i\mathbb{R}}\left|\mathrm{e}^{\eta_m T-2\gamma(s)\kappa(s,\lambda_m)}\right|\,\|\gamma_0 H\|_{L^2(0,T;L^2(\Sigma))},$$

where C depends polynomially on T, cf. this expression with (27). It remains to find suitable parameters $\eta_m>0$ to minimize the supremum on the right hand side (uniformly in λ_m).

Comparing the two techniques described above one notices that in both cases one has to find a contour $\mathcal{C}_m$ (in the latter case a straight line) in $\mathbb{C}_*^+$ connecting $-i\infty+\eta_-$ and $i\infty+\eta_+$, η_-, $\eta_+>0$, such that the quantity

$$A(s,\lambda_m,t)=\mathrm{Re}\left(st-2\gamma(s)\kappa(s,\lambda_m)\right),\quad s\in\mathcal{C}_m, \tag{28}$$

can be bounded on this contour by a negative number, so that $|\mathrm{e}^{-\mathrm{Re}A(s,\lambda_m,t)}|$ is small. For convergence, it is necessary that this bound decreases with $\bar\sigma$, L (but, possibly, deteriorates with t). Also this bound has to be uniform in λ_m.

When using the first method one may use more general contours, while when using Plancherel's identity one is limited to straight, vertical lines. Thus, a priori the first technique might lead to a better bound. However, as we shall see in the following Subsection 3.1, a properly chosen straight line leads to error bounds which are in some sense optimal. Therefore, we will present in Subsection 3.2 the proof of Theorem 1.12 using Plancherel's identity.

3.1. **On obtaining the error bounds by deforming the Laplace inversion contour**

Due to the considerations above we have to find a contour $\mathcal{C}_m$ in the complex right half plane connecting $-i\infty + \eta_-, i\infty + \eta_+$ for some $\eta_+, \eta_- > 0$ such that

$$A(s, \lambda_m, t) \leq \beta_m < 0, \quad s \in \mathcal{C}_m, \tag{29}$$

with A given by (28) and some $\beta_m < 0$.

First, we will show how to do so in the non-dispersive case, next discuss the dispersive case, and finally argue that the bounds obtained using this approach, in general, do not improve over the use of the usual Plancherel bound, cf. [BK21].

Let us emphasize that the approach presented here is different from the one presented in [BK21] for the non-dispersive case, and the goal of this small section is actually to show that, in general, a naively more optimal approach would not necessarily lead to better uniform error estimates.

3.1.1. *Non-dispersive case*

For the remainder of this subsection we assume that $\mu = \varepsilon \equiv 1$ and $t > 2L$, i.e.

$$A(s, \lambda_m, t) = \operatorname{Re}\left(st - 2L\left(1 + \tfrac{\bar{\sigma}}{s}\right)\sqrt{\lambda_m^2 + s^2}\right). \tag{30}$$

To obtain the best bound on the error we want to find an integration contour $\mathcal{C}_m$, such that the constant β_m in (29) is minimized. Remark that such a contour should lie in $\mathbb{C}_*^+$, and, moreover, connect $\eta_- - i\infty$ and $\eta_+ + i\infty$, $\eta_\pm > 0$.

The principal result of this section reads.

Theorem 3.1. *Let $\lambda_m > 0$, $t > 2L$, and $\bar{\sigma}$ be fixed. Then there exists an optimal contour $\mathcal{C}_m$ (in a sense that β_m in (29) is minimized), which passes through the two unique saddle points $s_0, \bar{s}_0$ of $A(\cdot, \lambda_m, t)$ in $\mathbb{C}_*^+$ and connects $\frac{A(s_0,\lambda_m,t)+2L\bar{\sigma}}{t-2L} - i\infty$ and $\frac{A(s_0,\lambda_m,t)+2L\bar{\sigma}}{t-2L} + i\infty$. Moreover, one can choose this contour so that $A(s, \lambda_m, t) = A(s_0, \lambda_m, t) = \beta_m \in [-2L\bar{\sigma}, 0)$ for all $s \in \mathcal{C}_m$.*

Remark 3.2. The case $\lambda_m = 0$ is quite simple: $A(s, 0, t) = (t - 2L)\mathrm{Re}s - 2L\bar{\sigma}$, and the value of $A(s, 0, t)$ is constant along the lines $i\mathbb{R} + \eta$, $\eta > 0$, and is decreasing in $\mathrm{Re}s$.

The proof of the above result relies on the following two technical propositions, which we prove in Appendices B and C. The first proposition describes the behaviour of the level sets of real parts of holomorphic functions.

Proposition 3.3. *Let $D \subseteq \mathbb{C}$ be a possibly unbounded domain, and ϕ be non-constant and holomorphic in D. Let ϕ have a finite number of stationary points. For $\phi_R = \operatorname{Re}\phi$, it holds that*

1. *the level sets of ϕ_R are finite unions of piecewise-smooth curves, which may intersect only in stationary points of ϕ;*
2. *the level sets of ϕ_R do not contain closed contours;*
3. *each component of a level set of ϕ_R approaches ∂D or is unbounded;*
4. *each stationary point of ϕ_R is a point of a self-intersection of the level set; moreover, the branches of the level set curves intersecting in s_0 divide the vicinity $B_\varepsilon(s_0)$ for $\varepsilon \to 0$ into $2 + 2m$ sectors where m is the order of the stationary point.*

The second proposition shows that $P(s) := st - 2\gamma(s)\kappa(s, \lambda_m)$ has only two saddle points in $\mathbb{C}_*^+$, cf. (30).

Proposition 3.4. *For all $\bar{\sigma} > 0$, $L > 0$, $t > 2L$, the function $s \mapsto (st - 2\gamma(s)\kappa(s, \lambda_m))'$ has precisely two complex conjugate roots with a positive real part.*

These two propositions allow us to prove Theorem 3.1.

Proof of Theorem 3.1. Our idea is to look for the contour $\mathcal{C}_m$ as a component of the level set $A(s, \lambda_m, t) = \beta_m$.

The function $s \mapsto A(s, \lambda_m, t)$ is the real part of the function $P(s) = st - 2\gamma(s)\kappa(s, \lambda_m)$. The function P is holomorphic in $\mathbb{C}_*^+$ and can be extended by continuity, together with its first derivative, to $\mathbb{C}^+ \setminus \{0, \pm i\lambda_m\}$.

By Proposition 3.4 we know that P has precisely two complex conjugate stationary points s_m^0, $\overline{s_m^0}$ with a positive real part. Since $A = \mathrm{Re}P$, the points s_m^0, $\overline{s_m^0}$ are unique stationary points of $s \mapsto A(s, \lambda_m, t)$ in $\mathbb{C}_*^+$. We assume without loss of generality that $\mathrm{Im}(s_m^0) > 0$.

Step 1. Properties of a level set passing through a saddle point. Our goal is to show that there exists a piecewise-smooth curve passing through s_m^0, s.t. $A(s, \lambda_m, t) = \mathrm{const} < 0$ along this curve; this curve would lie in $\mathbb{C}_*^+$, and connect the real axis to $\frac{A(s_m^0, \lambda_m, t) + 2L\bar{\sigma}}{t - 2L} + i\infty$. Note that due to symmetry of the function A with respect to the complex conjugation this is sufficient.

Step 1.1. Behaviour of A in $\mathbb{C}_*^+$.

Step 1.1.1. Behaviour of A on the real axis. Studying

$$D(s) := P'(s) = t - \frac{2sL}{\sqrt{s^2 + \lambda_m^2}} \left(1 + \frac{\bar{\sigma}}{s}\right) + \frac{2\bar{\sigma}L}{s^2} \sqrt{s^2 + \lambda_m^2} = t - \frac{2Ls}{\sqrt{s^2 + \lambda_m^2}} + \frac{2L\bar{\sigma}\lambda_m^2}{s^2 \sqrt{s^2 + \lambda_m^2}}, \tag{31}$$

we immediately obtain (using $t > 2L$) that $D(r) > t - 2L > 0$ for $r > 0$. Since

$$\lim_{r \to 0^+} A(r, \lambda_m, t) = -\infty, \quad \lim_{r \to +\infty} A(r, \lambda_m, t) = +\infty,$$

we have that the mapping $A|_{\mathbb{R}_*^+} \to \mathbb{R}$ is bijective.

Step 1.1.2. Behaviour of A on the imaginary axis.

Remark that the continuation of $D : \mathbb{C}_*^+ \to \mathbb{C}$ on $i\mathbb{R}$ satisfies $\mathrm{Im}(D(i\omega)) < 0$ for $\omega \in (0, \lambda_m)$ and $\mathrm{Im}(D(i\omega)) > 0$ for $\omega > \lambda_m$. Thus, using the limits

$$\lim_{\omega \to 0^+} A(i\omega, \lambda_m, t) = -2L\lambda_m, \qquad \lim_{\omega \to +\infty} A(i\omega, \lambda_m, t) = -2L\bar{\sigma},$$

and the fact that $A(i\lambda_m, \lambda_m, t) = 0$ we have that the mappings

$$A|_{s \in i(0, \lambda_m)} \to (-2L\lambda_m, 0), \qquad A|_{s \in i(\lambda_m, \infty)} \to (-2L\bar{\sigma}, 0), \tag{32}$$

are bijections.

Moreover, in the vicinity of $i\lambda_m$ in $\mathbb{C}_*^+$, we have that

$$A(i\lambda_m + \delta, \lambda_m, t) = -2L\sqrt{\lambda_m} \left(1 + \frac{\bar{\sigma}}{\lambda_m}\right) \delta^{1/2} + O(\delta), \quad \delta \to 0.$$

This shows in particular that the level sets in $\mathbb{C}^+$ along which $A = 0$ do not pass through $i\lambda_m$.

Step 1.1.3. Behaviour of A on the lines $r + i\omega$, $r > 0$, $|\omega| \gg 1$. We see that as $|s| \to +\infty$,

$$D(s) = t - 2L + O(1/|s|^3), \quad A(s, \lambda_m, t) = \mathrm{Re}s(t - 2L) - 2L\bar{\sigma} + O(1/|s|), \tag{33}$$

and therefore, on the lines $r + i\omega$, $r \geq 0$, with ω being a large constant, the function $r \mapsto A(r + i\omega, \lambda_m, t)$ is strictly monotonically increasing, from $-2L\bar{\sigma} + o(1)$ to $+\infty$.

Step 1.1.4. Behaviour of A on the lines $r + i\omega$, $r \gg 1$, $\omega \in \mathbb{R}$. By the same argument as before, on the lines $r + i\omega$, $\omega \in \mathbb{R}$, with r being a large constant, the function $A(r + i\omega, \lambda_m, t) = r(t - 2L) + o(1)$.

Step 1.1.5. Behavior of A at $s = 0$. Expanding A in the series in the vicinity of the origin yields

$$A(s, \lambda_m, t) = -2L\lambda_m \frac{\bar{\sigma}}{|s|^2} \mathrm{Re}s - 2L\lambda_m + O(|s|), \quad s \to 0,$$

which gives $A(s, \lambda_m, t) < -2L\lambda_m$ for $s \in \mathbb{C}_*^+$ sufficiently close to 0.

These results are summarized in Figure 2.

Step 1.2. Proof that $A(s_m^0, \lambda_m, t) \in (-2L\bar{\sigma}, 0)$. Let us set now $\beta_m := A(s_m^0, \lambda_m, t)$. Then we know by Proposition 3.3 that in s_m^0 the level set self-intersects, and there are at least 4 different curves originating at s_m^0, along which $A(\cdot, \lambda_m, t)$ is constant, and which do not intersect but in s_m^0 and, possibly, in $\overline{s_m^0}$.

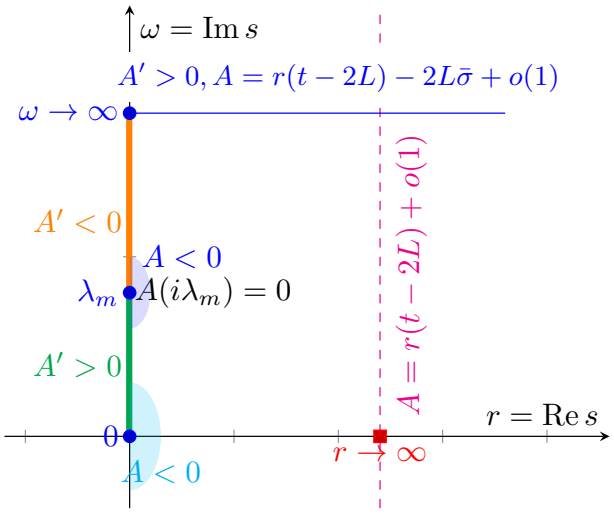


FIGURE 2. Behaviour of A in the right-half complex plane

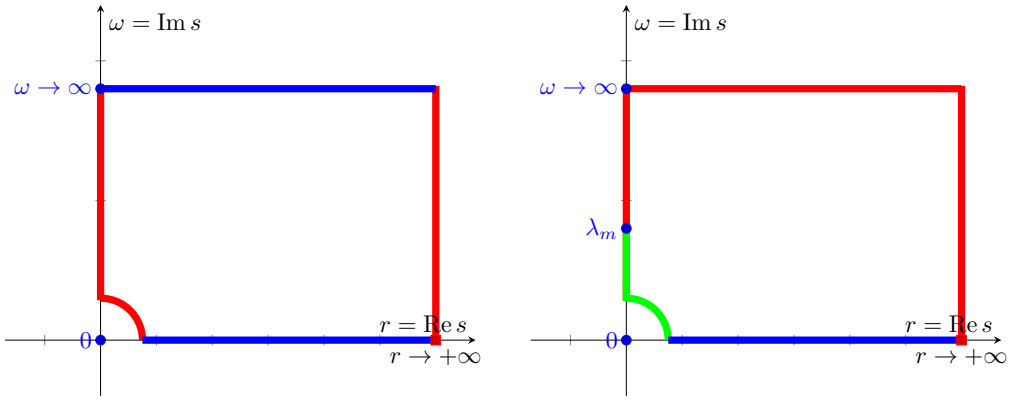


FIGURE 3. 'Forbidden' components of the boundary of the domain $D_{\varepsilon,R}$, for ε sufficiently small and R sufficiently large. Left: the components of boundaries of $D_{\varepsilon,R}$ that cannot be crossed by level sets $A(s, \lambda_m, t) = \text{const} > 0$ are marked in red, and those that can be crossed at most once in blue. Right: the components of boundaries of $D_{\varepsilon,R}$ that cannot be crossed by level sets $A(s, \lambda_m, t) = \text{const} \leq -2L\bar{\sigma}$ are marked in red, and those that can be crossed at most once in blue (resp. green).

First of all, by Proposition 3.3, these curves should intersect the boundary of the domain

$$D_{\varepsilon,R} = \{z \in \mathbb{C}^+_*, \quad \varepsilon < \operatorname{Im} z, \operatorname{Re} z < R, \quad |z| > \varepsilon\},$$

for any $\varepsilon > 0$, $R > 0$. However, because of the monotonicity properties of the function A discussed above, we have that, for sufficiently large R and sufficiently small $\varepsilon > 0$,

- the level set $A(s^0_m, \lambda_m, t) = \beta_m$ intersects the real axis at most once;
- for sufficiently large R it does not intersect the boundary along which $\operatorname{Re} z = R$;
- the imaginary axis and $\partial B_\varepsilon(0) \cap \partial D$ can be intersected only if $\beta_m < 0$ (remark in particular the behaviour of A in the vicinity of λ_m and 0).

This immediately implies that $\beta_m < 0$, cf. Figure 3, left.

Next, let us show that $\beta_m > -2L\bar{\sigma}$. As follows from the behaviour of the function $s \mapsto A(s, \lambda_m, t)$ in $\mathbb{C}^+_*$, the level set on which $\beta_m \leq -2L\bar{\sigma}$

- does not intersect the boundaries $\mathrm{Im} z = R$, $\mathrm{Re} z = R$ for R sufficiently large;
- intersects the real axis at most once;
- intersects at most once $(\partial B_\varepsilon(0) \cap \partial D_{\varepsilon,R}) \cup (i\varepsilon, i\lambda_m)$ (in particular, if $\beta_m \leq -2L\bar{\sigma}$ and $-2L\bar{\sigma} > -2L\lambda_m$, this level set would intersect $(i\varepsilon, i\lambda_m)$, for $-2L\bar{\sigma} \leq -2L\lambda_m$, it would intersect $\partial B_\varepsilon(0) \cap \partial D_{\varepsilon,R}$).

This is illustrated in Figure 3, right. Because in total in $D_{\varepsilon,R}$ there are at least four non-intersecting curves emanating from s^0_m, we arrive at a conclusion that $\beta_m > -2L\bar{\sigma}$.

Step 2. Construction of the contour $\mathcal{C}_m$. Let us now prove that there exists a curve connecting $z_{r,0} + i\infty$ the saddle point s^0_m and a point on $\mathbb{R}^*_+$, s.t. $z_{r,0} > 0$ and A is constant along this curve.

Let us consider the branches of the level set curve starting at the point s^0_m. Because $\beta_m \in [-2L\bar{\sigma}, 0)$, by (32), at most two of these branches γ_1, γ_2 intersect $(i\mathbb{R} \cup \partial B_\varepsilon(0)) \cap \partial D_{\varepsilon,R}$: they intersect correspondingly the interval $i(\lambda_m, +\infty)$ and either the interval $(0, i\lambda_m)$ or $\partial B_\varepsilon(0) \cap \partial D_{\varepsilon,R}$ for sufficiently small $\varepsilon > 0$. For the remaining branches the following holds true: either the branch is unbounded in $\mathbb{C}^+ \cap \{\mathrm{Im} z > 0\}$ or it intersects the real axis. However, by studying the behaviour of A at infinity (and its monotonicity properties), we see that only a single branch can be unbounded, and only one branch can intersect the real axis. The unbounded branch $\mathcal{C}^{+,\infty}_m$ necessarily intersects all the lines $\{r + i\omega,\, r > 0\}$, for all ω sufficiently large (since $A < 0$ along this branch). By (33), this branch connects $z_{r,0} + i\infty$ with $z_{r,0}$ as in the statement of the theorem, and s^0_m. The remaining branch $\mathcal{C}^{+,0}_m$ then intersects the real axis.

Denoting by $\mathcal{C}^{-,\infty}_m = \{\bar{z},\, z \in \mathbb{C}^{+,\infty}_m\}$, and using a similar convention for $\mathcal{C}^{-,0}_m$, we obtain

$$\mathcal{C}_m = \overline{\mathcal{C}^{-,\infty}_m \cup \mathcal{C}^{+,\infty}_m \cup \mathcal{C}^{-,0}_m \cup \mathcal{C}^{+,0}_m}.$$

Step 3. Optimality of the contour.

It remains to show that the chosen curve is optimal. Assume that there exists another contour $\mathcal{C}'_m$ passing through the real axis and joining $\eta_- + i\infty$ and $\eta_+ + i\infty$, with $\eta_\pm \geq 0$, along which $A(s, \lambda_m, t) \leq \beta_m - \varepsilon$, for some $\varepsilon > 0$. Let us denote the point where the contour $\mathcal{C}'_m$ crosses the real axis by x'_0, and the point where the contour $\mathcal{C}_m$ crosses the real axis by x_0. By monotonicity of A on the real axis, $x'_0 < x_0$. From the considerations in Step 2 we see that there is a curve passing through s^0_m and joining a point $(x, 0)$ on the real axis and some point on the imaginary axis, along which the function A equals to β_m. Because $x'_0 < x_0$, any contour joining x'_0 to $\eta_+ + i\infty$, $\eta_+ > 0$, would cross this curve. However, on this curve $A(s, \lambda_m, t) = \beta_m$, and thus we have arrived at a contradiction with our assumption. □

Let us provide several comments about the result of the above theorem.

Non-uniqueness of the contour. The theorem shows that an optimal contour exists, but of course such a contour is non-unique. As we see in Figure 4, we could have taken instead of a level set a straight line passing through s^0_m.

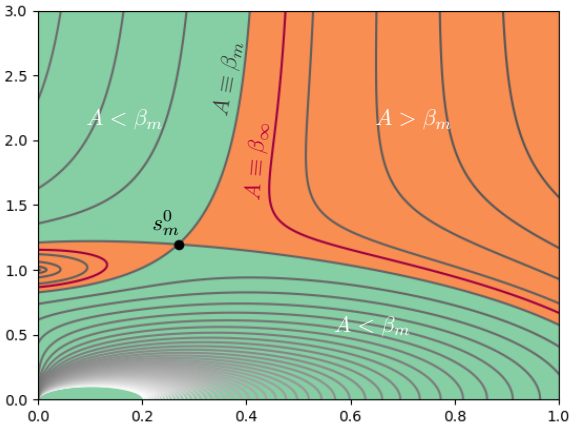
Optimality of the error and the use of the result of the theorem for an a priori error estimate. Moreover, the proof of the theorem is constructive: it allows to find an optimal bound on $A(s, \lambda_m, t)$, which, in turn, gives quite a good indication of the behaviour of the PML error for each mode λ_m. In particular, it says that the optimal value β_m is given by $A(s^0_m, \lambda_m, t)$, and s^0_m can be found numerically as a root of the equation $D(s) = 0$. However, for the moment we did not provide any analytic bounds on $A(s^0_m, \lambda_m, t)$.

As for deriving a uniform in λ_m bound, it appears that as $\lambda_m \to +\infty$, the quantity β_m converges. Then a uniform in λ_m error bound cannot be better than the one given by the limit of β_m.

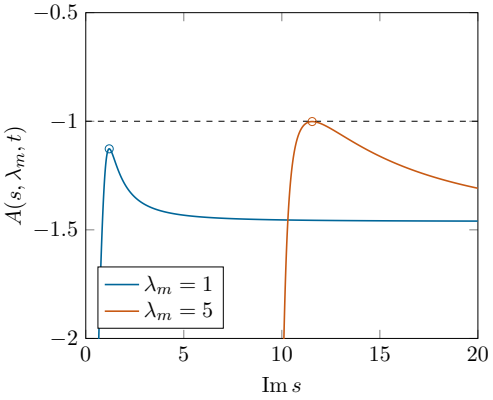
Proposition 3.5. *Let β_m and s^0_m be as in Theorem 3.1; $t > 0$. Then, as $\lambda_m \to +\infty$,*

$$s^0_m = i\lambda_m \sqrt{\frac{t^2}{t^2 - 4L^2}} + \frac{4L^2\bar{\sigma}}{t^2} + O(\lambda_m^{-1}), \quad \beta_m = \beta_\infty + O\big(\lambda_m^{-1}\big), \quad \beta_\infty = -\frac{4L^2\bar{\sigma}}{t}.$$

Proof. It suffices to find the roots of $D(s)$ defined in (31) with a positive real part for $\lambda_m \to +\infty$ (recall that by Proposition 3.4, there is one complex-conjugate pair of roots in $\mathbb{C}^+_*$). For this we rescale $s_m := \lambda_m s$,



(A) level sets, stationary and paths of integration for the function $A(\cdot, \lambda_m, t)$ for $\lambda_m = 1$



(B) $A(s, \lambda_m, t)$ along the lines $\mathrm{Re}\, s = \mathrm{Re}\, s_m^0$ for two different values of λ_m.

FIGURE 4. Behavior of the function $A(s, \lambda_m, t)$ for $\bar{\sigma} = L = 1, t = 4$ in the non-dispersive case (see Proposition 3.5 for the notation β_∞). The circles mark the (imaginary parts) of the saddle points and the dashed line the global bound $\beta_\infty = 1$.

which yields

$$D(s_m) = t - \frac{2Ls_m}{\sqrt{s_m^2 + 1}} + \frac{2L\bar{\sigma}}{\lambda_m s_m^2 \sqrt{s_m^2 + 1}}.$$

By the implicit function theorem, one of the roots of $D(s_m)$ is given by

$$s_m^* = i\sqrt{\frac{t^2}{t^2 - 4L^2}} + \frac{4L^2\bar{\sigma}}{t^2\lambda_m} + O(\lambda_m^{-2}).$$

Remarking that

$$\sqrt{(s_m^*)^2 + 1} = \frac{1}{t}\left(2Ls_m^* - \frac{2L\bar{\sigma}}{\lambda_m (s_m^*)^2}\right),$$

we can compute the expansion of β_m, as $\lambda_m \to +\infty$,

$$\beta_m = A(s_m^*/\lambda_m, \lambda_m, t) = -\frac{4L^2\bar{\sigma}}{t} + O(\lambda_m^{-1}).$$

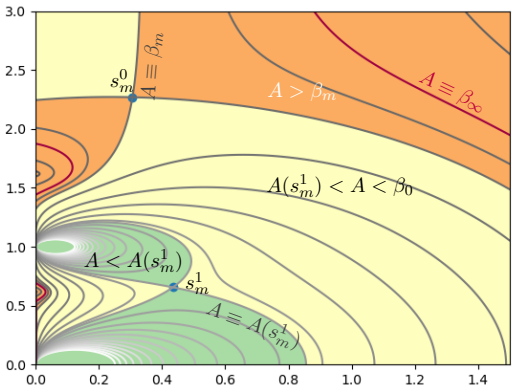
□

We will see in the section that follows that the above limit provides the best uniform in λ_m error bound.

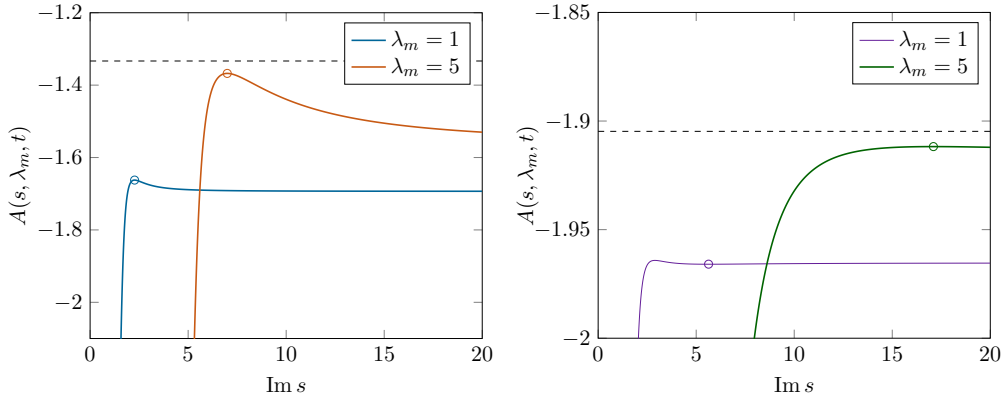
Remark 3.6. The fact that finding a stationary of $A(\cdot, \lambda_m, t)$ can be reduced to the root finding of $D(s) = 0$, where $D(s)$ is defined in (31), which, in turn can be recast as a sixth order polynomial equation, leads to a method to compute the optimal bounds numerically, in the case when the data is highly regular in space.

3.1.2. *Dispersive media: open questions*

For the dispersive media, the reasoning as above applies as well: we look at the contour where $A(\cdot, \lambda_m, t)$ is minimized as a sub-contour of a level set passing through a saddle point of $st - 2\gamma(s)\kappa(s, \lambda_m)$. However,



(A) Level sets, stationary points and paths of integration for the function $A(\cdot, \lambda_m, t)$ for $\lambda_m = 1$



(B) $A(s, \lambda_m, t)$ along the lines $\mathrm{Re}\, s = \mathrm{Re}\, s_m^0$ for two different values of λ_m and $t = 3$ (left) and $t = 2.1$ (right). The circles mark the (imaginary parts) of the saddle points and the dashed line the global bound β_∞.

FIGURE 5. Behavior of the function $A(s, \lambda_m, t)$ for $\bar{\sigma} = L = 1$ and Drude media with $\mu_0 = \varepsilon_0 = 1$. Remark that we use the notation $A(s)$ for $A(s, \lambda_m, t)$.

the difficulty in choosing this contour lies in the fact that for dispersive media more than one stationary point might exist, as illustrated in Figure 5.

On the other hand, because of the non-dispersivity at high frequencies (cf. Section 1.1.1) we can expect that the behaviour of $A(s, \lambda_m, t)$ for large $|s|$ in the dispersive case resembles the behaviour of $A(s, \lambda_m, t)$ in the non-dispersive case, and, therefore, one can conjecture that the best uniform error bound for A in this case is also given by the result of Proposition 3.5.

3.1.3. *Discussion*

The above numerical experiments and semi-analytical arguments indicate that

- we can hope to find a contour along which the function $A(s, \lambda_m, t)$ is uniformly bounded by looking at the contours with $\mathrm{Re}\, s = \text{const}$, cf. Figure 4 and 5. Remark however that, as Figure 5b illustrates, the maximum value of A along the straight line with $\mathrm{Re}\, s = \mathrm{Re}\, s_m^0$ is not necessarily given by $A(s_m^0, \lambda_m, t)$, contrary to what numerical experiments indicate in the non-dispersive case.
- for large λ_m, we expect that an upper bound for $A(s, \lambda_m, t)$ that would be close to the optimal one is given by $\beta_\infty = -4L^2\bar{\sigma}/t$, cf. Proposition 3.5 and the discussion in the end of Section 3.1.2. Moreover, the numerical experiments in Figure 4 and the argument of Proposition 3.5 indicate that

in this case a 'good' choice of the contour is given by $\mathrm{Re} s = \frac{4L^2\bar{\sigma}}{t^2}$. Let us remark that this choice is close to the one suggested in [BK21] for all λ_m ($\mathrm{Re} s = \frac{4L^2\bar{\sigma}}{t^2} c$, with $c = (1 + \frac{16L^2}{t^2})^{-1}$).

3.2. **Proof of Theorem 1.12**

We have seen above that choosing a contour along a level set $A(s, \lambda_m, t) = \beta_m < 0$ allows us to bound the error, cf. (27). The results of the previous section show that such a contour can be chosen for any λ_m and, moreover, for $\lambda_m \to +\infty$, $\beta_m \to \beta_\infty = -4L^2\bar{\sigma}^2 t^{-1}$. In this section we show that this latter bound is suitable for all λ_m, $m \in \mathbb{N}$; moreover, it can be also achieved by using the contour $\mathcal{C} = \{s \in \mathbb{C} : \mathrm{Re} s = \eta\}$ for a suitable $\eta > 0$, independent of m. We conclude this subsection by filling in the remaining details of the proof of Theorem 1.12 where we make use of Plancherel's Theorem.

3.2.1. *Several auxiliary technical results*

We start with the following result about $\kappa(s, \lambda) = \sqrt{s^2\varepsilon(s)\mu(s) + \lambda^2}$.

Lemma 3.7. *For $\lambda \in \mathbb{R}$, the function $s \mapsto \kappa(s, \lambda)$ is analytic in $\mathbb{C}_*^+$. Moreover,*

$$\mathrm{Re}\, \kappa(s, \lambda) > \mathrm{Re} s.$$

Proof. Analyticity. The analyticity of $s \mapsto \kappa(s, \lambda_m) = \sqrt{s^2\varepsilon(s)\mu(s) + \lambda_m^2}$ in $\mathbb{C}_*^+$ follows from the passivity assumption (3): indeed, $\mathrm{Re}(s\varepsilon(s)) > 0$, $\mathrm{Re}(s\mu(s)) > 0$ in $\mathbb{C}_*^+$, therefore $\mathrm{Arg}(s^2\varepsilon(s)\mu(s)) \in (-\pi, \pi)$, and thus $\mathrm{Arg}(s^2\varepsilon(s)\mu(s) + \lambda_m^2) \in (-\pi, \pi)$.

Next, to prove a lower bound on κ, let us recall that for $z \in \mathbb{C} \setminus \mathbb{R}_{\le 0}$, the principal square root is

$$\sqrt{z} = \frac{1}{\sqrt{2}} \left(\sqrt{|z| + \mathrm{Re}(z)} + \mathrm{sign}(\mathrm{Im}(z))\, i \sqrt{|z| - \mathrm{Re}(z)} \right). \tag{34}$$

Hence $x \mapsto \mathrm{Re}\sqrt{x + iy}$ is non-decreasing in $x \in \mathbb{R}$. Hence

$$\mathrm{Re}\kappa(s, \lambda) \ge \mathrm{Re}\kappa(s, 0) = \mathrm{Re}\sqrt{s^2\mu(s)\varepsilon(s)}.$$

From (34) it follows that, for $z_k = z_{kr} + iz_{ki}$, $k = 1, 2$,

$$\begin{aligned}(\mathrm{Re}\sqrt{z_1 z_2})^2 = \frac{1}{2}(|z_1 z_2| + z_{1r}z_{2r} - z_{1i}z_{2i}) &= \frac{1}{2}(\sqrt{z_{1r}^2 + z_{1i}^2}\sqrt{z_{2r}^2 + z_{2i}^2} + z_{1r}z_{2r} - z_{1i}z_{2i}) \\ &\ge \frac{1}{2}(|z_{1r}z_{2r}| + |z_{1i}z_{2i}| + z_{1r}z_{2r} - z_{1i}z_{2i}),\end{aligned}$$

where the last inequality was obtained from the Cauchy-Schwarz inequality. Applying the above with $z_1 = s\mu(s)$ and $z_2 = s\varepsilon(s)$ yields

$$\mathrm{Re}\kappa(s, 0) \ge \sqrt{\mathrm{Re}(s\mu(s))\mathrm{Re}(s\varepsilon(s))} \ge \mathrm{Re} s,$$

where the last inequality follows from the passivity condition (3). □

We will also need the following techical result (whose role will be clear later).

Lemma 3.8. *For all $s \in \mathbb{C}_*^+$ and $\lambda \in \mathbb{R}$*

$$\mathrm{Re}\kappa(s, \lambda)\, \mathrm{Re}\left(\frac{\kappa(s, \lambda)}{s\zeta(s)} \right) \ge \mathrm{Re} s. \tag{35}$$

Proof. Let $s \in \mathbb{C}_*^+$. For brevity of notation, we write κ for $\kappa(s,\lambda)$ and ζ for $\zeta(s)$ (resp. μ, ε). We rewrite the left hand side of (35) by replacing $\mathrm{Re}\kappa$ from equation (34), and recalling that $\mathrm{Im}\kappa^2 = 2\,\mathrm{Re}\kappa\,\mathrm{Im}\kappa$

$$\begin{aligned}
\mathrm{Re}\kappa\,\mathrm{Re}\frac{\kappa}{s\zeta} =& \frac{\mathrm{Re}\kappa}{|s\zeta|^2}\left(\mathrm{Re}\kappa\,\mathrm{Re}(s\zeta) + \mathrm{Im}\kappa\,\mathrm{Im}(s\zeta)\right)\\
=& \frac{1}{2|s\zeta|^2}\left(\left(|\kappa|^2 + \mathrm{Re}\big(\kappa^2\big)\right)\mathrm{Re}(s\zeta) + \mathrm{Im}\big(\kappa^2\big)\,\mathrm{Im}(s\zeta)\right)\\
=& \frac{1}{2|s\zeta|^2}\mathrm{Re}(s\zeta)\left(|s^2\varepsilon\mu + \lambda^2| + \mathrm{Re}\big(s^2\varepsilon\mu + \lambda^2\big)\right) + \frac{1}{2|s\zeta|^2}\mathrm{Im}(s\zeta)\,\mathrm{Im}\big(s^2\varepsilon\mu + \lambda^2\big)\,.
\end{aligned}$$

Since $\mathrm{Re}(s\zeta) > 0$ by the passivity of ζ we use the inverse triangle inequality for the first term in the above

$$|s^2\varepsilon\mu + \lambda^2| + \mathrm{Re}\big(s^2\varepsilon\mu + \lambda^2\big) \geq |s^2\varepsilon\mu| + \mathrm{Re}\big(s^2\varepsilon\mu\big)$$

to obtain

$$\begin{aligned}
\mathrm{Re}\kappa\,\mathrm{Re}\frac{\kappa}{s\zeta} &\geq \frac{1}{2|s\zeta|^2}\left[\mathrm{Re}(s\zeta)\left(|s^2\varepsilon\mu| + \mathrm{Re}\big(s^2\varepsilon\mu\big)\right) + \mathrm{Im}(s\zeta)\,\mathrm{Im}\big(s^2\varepsilon\mu\big)\right]\\
&= \frac{1}{2}\left[\frac{|\varepsilon\mu|\mathrm{Re}(s\zeta)}{|\zeta|^2} + \frac{\mathrm{Re}\big(s^2\varepsilon\mu\big)\,\mathrm{Re}(s\zeta) + \mathrm{Im}(s\zeta)\,\mathrm{Im}\big(s^2\varepsilon\mu\big)}{|s\zeta|^2}\right]\\
&= \frac{1}{2}\left[\frac{|\varepsilon\mu|\mathrm{Re}(s\zeta)}{|\zeta|^2} + \frac{\mathrm{Re}\big(\overline{s\zeta}s^2\varepsilon\mu\big)}{|s\zeta|^2}\right] = \frac{1}{2}\left[\frac{|\varepsilon\mu|\mathrm{Re}(s\zeta)}{|\zeta|^2} + \mathrm{Re}\left(\frac{s\varepsilon\mu}{\zeta}\right)\right] \quad \geq \frac{1}{2}\left(\mathrm{Re}s\left|\frac{\mu\varepsilon}{\zeta}\right| + \mathrm{Re}\left(\frac{s\mu\varepsilon}{\zeta}\right)\right),
\end{aligned}$$

where the last bound follows from (MP) in the form $\frac{\mathrm{Re}(s\zeta)}{|\zeta|} \geq \mathrm{Re}s$.

Finally, applying again (MP) and next using the inequality $a + \max(a, 2-a) \geq 2$ yields the desired bound:

$$\mathrm{Re}\kappa\,\mathrm{Re}\frac{\kappa}{s\zeta} \geq \frac{1}{2}\mathrm{Re}s\left(\left|\frac{\mu\varepsilon}{\zeta}\right| + \max\left(\left|\frac{\mu\varepsilon}{\zeta}\right|, 2 - \left|\frac{\mu\varepsilon}{\zeta}\right|\right)\right) \geq \mathrm{Re}s.$$

□

3.2.2. *Proof of Theorem 1.12*

The two Lemmas 3.7 and 3.8 allow us to prove the principal proposition about the choice of the contour that allows to control the quantity (28).

Proposition 3.9. *Let μ, ε, ζ such that Assumptions 1.2 and 1.11 hold. Then for given $\lambda_m \in \mathbb{R}$ and $t, L, \bar{\sigma} \geq 0$ we have that that the function A defined in* (28) *is holomorphic in $\mathbb{C}_*^+$ and fulfills*

$$A\left(\frac{4L^2\bar{\sigma}}{t^2} + is_i, \lambda_m, t\right) \leq -\frac{4L^2\bar{\sigma}}{t}, \quad s_i \in \mathbb{R}. \tag{36}$$

Before proving this proposition, we remark that the quantity A controlling the error is uniformly bounded in λ_m on a straight line $\frac{4L^2\bar{\sigma}}{t^2} + is_i$. According to Proposition 3.5, as $\lambda_m \to +\infty$, in the non-dispersive case, this line gets closer and closer to the straight line passing through the saddle point s_m^0 (see previous section for the discussion of the role of the saddle points in the error bounds). Moreover, the upper bound is consistent with the best upper bound of Proposition 3.5.

Proof of Proposition 3.9. Recall that A is given by

$$A(s,\lambda,t) = \mathrm{Re}(st) - \left(2L\mathrm{Re}\kappa(s,\lambda) + 2L\bar{\sigma}\mathrm{Re}\left(\frac{\kappa(s,\lambda)}{s\zeta(s)}\right)\right).$$

By Lemma 3.7 $A(\cdot,\lambda,t)$ is holomorphic in $\mathbb{C}_*^+$. We next apply the Young inequality $a+b\geq 2\sqrt{ab}$ with

$$a=2L\mathrm{Re}\kappa(s,\lambda), \qquad\qquad b=2L\bar{\sigma}\mathrm{Re}\left(\frac{\kappa(s,\lambda)}{s\zeta(s)}\right),$$

and use the bound of Lemma 3.8 for $\sqrt{ab}$ to obtain

$$A(s,\lambda,t)\leq t\mathrm{Re}s-4L\sqrt{\bar{\sigma}\mathrm{Re}s}.$$

This shows that on straight vertical lines A can be bounded independently of $\mathrm{Im}s$ and λ. An 'optimal' line can be found by minimizing A in $\mathrm{Re}s>0$. In particular, we remark that the function $g(x)=tx-4L\sqrt{\bar{\sigma}}\sqrt{x}$, takes its global minimum at $x_0=\frac{4L^2\bar{\sigma}}{t^2}$, and this minimum is equal to $g(x_0)=-\frac{4L^2\bar{\sigma}}{t}$. Hence the result in the statement of the proposition follows. □

Now we are in position to prove Theorem 1.12.

Proof of Theorem 1.12. First of all, let us remark that the PML problem admits a unique solution, as per Corollary 2.4. The proof of Theorem 1.12 follows like in the proof of Theorem 4.1 from [BK21]. We will use the notation $a\lesssim b$ to indicate that $a\leq Cb$, for a generic constant $C>0$ independent of t, s and the parameters of the problem.

Case $t>2L$. Using Plancherel's theorem, for $\eta>0$ one has that, cf. (26),

$$\int_0^\infty \mathrm{e}^{-2\eta t}\|e_m^\sigma(t)\|^2_{L^2(\Omega_{ph})}dt=\frac{1}{2\pi}\int_{\eta+i\mathbb{R}}\mathrm{e}^{-4\mathrm{Re}(\kappa(s,\lambda_m)\gamma(s))}\|c_m(s,.)\|^2_{L^2(0,R)}ds. \tag{37}$$

A straightforward computation yields the following bound:

$$\begin{aligned}\|c_m(s,.)\|^2_{L^2(0,R)}&\leq|1-\mathrm{e}^{-2\kappa(s,\lambda_m)(\gamma(s)+R)}|^{-2}\frac{1}{|\kappa(s,\lambda_m)|^2}\left(1+|\mathrm{e}^{-2\kappa(s,\lambda_m)R}|\right)|\hat{H}_m(s,R)|^2\\&\lesssim(\mathrm{Re}s)^{-2}|1-\mathrm{e}^{-2\kappa(s,\lambda_m)(\gamma(s)+R)}|^{-2}|\hat{H}_m(s,R)|^2.\end{aligned} \tag{38}$$

Inserting this bound into (37) and using the Plancherel identity again yields

$$\int_0^\infty \mathrm{e}^{-2\eta t}\|e_m^\sigma(t)\|^2_{L^2(0,R)}dt\lesssim\eta^{-2}\sup_{s\in\eta+i\mathbb{R}}\left(\mathrm{e}^{-4\mathrm{Re}(\kappa(s,\lambda_m)\gamma(s))}|1-\mathrm{e}^{-2\kappa(s,\lambda_m)(\gamma(s)+R)}|^{-2}\right)\int_0^\infty\mathrm{e}^{-2\eta t}|H_m(t,R)|^2dt.$$

Next, one can use the causality argument (cf. [BK21]) which allows to truncate the above integrals to finite intervals. More precisely, for each $T>0$, it holds that

$$\int_0^T \mathrm{e}^{-2\eta t}\|e_m^\sigma(t)\|^2_{L^2(0,R)}dt\lesssim\eta^{-2}\sup_{s\in\eta+i\mathbb{R}}\left(\mathrm{e}^{-4\mathrm{Re}(\kappa(s,\lambda_m)\gamma(s))}|1-\mathrm{e}^{-2\kappa(s,\lambda_m)(\gamma(s)+R)}|^{-2}\right)\int_0^T\mathrm{e}^{-2\eta t}|H_m(t,R)|^2dt.$$

Finally, we bound the above integrals to obtain

$$\|e_m^\sigma\|^2_{L^2(0,T;L^2(0,R))}\lesssim\eta^{-2}\sup_{s\in\eta+i\mathbb{R}}\left(\mathrm{e}^{2\eta T-4\mathrm{Re}(\kappa(s,\lambda_m)\gamma(s))}|1-\mathrm{e}^{-2\kappa(s,\lambda_m)(\gamma(s)+R)}|^{-2}\right)\|H_m(.,R)\|^2_{L^2(0,T)}.$$

Repeating the above arguments with e_m^σ replaced by $\partial_t e_m^\sigma$, and using the fact that $\partial_t H_m(0,R)=0$, we get the same bound for $\|\partial_t e_m^\sigma\|^2_{L^2(0,T;L^2(0,R))}$. Finally, since $e_m^\sigma(0)=0$, we have that $\|e_m^\sigma\|_{L^\infty(0,T;L^2(0,R))}\leq T^{1/2}\|\partial_t e_m^\sigma\|_{L^\infty(0,T;L^2(0,R))}$. Moreover, $\gamma_0^\Sigma H\big|_{t=0}=0$, and therefore

$$\begin{aligned}\|e_m^\sigma\|^2_{L^\infty(0,T;L^2(0,R))}&\lesssim T\eta^{-2}\sup_{s\in\eta+i\mathbb{R}}\left(\mathrm{e}^{2\eta T-4\mathrm{Re}(\kappa(s,\lambda_m)\gamma(s))}|1-\mathrm{e}^{-2\kappa(s,\lambda_m)(\gamma(s)+R)}|^{-2}\right)\|\partial_t H_m(.,R)\|^2_{L^2(0,T)}\\&\lesssim T\eta^{-2}\sup_{s\in\eta+i\mathbb{R}}\exp(2A(s,\lambda_m,T))\sup_{s\in\eta+i\mathbb{R}}|1-\mathrm{e}^{-2\kappa(s,\lambda_m)(\gamma(s)+R)}|^{-2}\|\partial_t H_m(.,R)\|^2_{L^2(0,T)}.\end{aligned} \tag{39}$$

Since by Proposition 3.9 we are able to bound the first exponent in the supremum by setting $\eta = 4L^2\bar{\sigma}T^{-2}$, it remains to bound the term $|1 - \mathrm{e}^{-2\kappa(s,\lambda_m)(\gamma(s)+R)}|^{-2}$. We use

$$\left|1 - \mathrm{e}^{-2\kappa(s,\lambda_m)(\gamma(s)+R)}\right| \geq 1 - \mathrm{e}^{-2\mathrm{Re}(\kappa(s,\lambda_m)(\gamma(s)+R))}$$

and the fact that $A(s,\lambda,T) = \mathrm{Re}sT - 2\mathrm{Re}(\kappa(s,\lambda)\gamma(s,\lambda))$, as well as Lemma 3.7, which allows to rewrite for $\mathrm{Re}s = \eta$,

$$\left|1 - \mathrm{e}^{-2\kappa(s,\lambda_m)(\gamma(s)+R)}\right| \geq 1 - \mathrm{e}^{A(s,\lambda_m,T)-\mathrm{Re}sT-2\mathrm{Re}sR} \geq 1 - \mathrm{e}^{-8L^2\bar{\sigma}T^{-1}-8L^2RT^{-2}\bar{\sigma}}.$$

With the inequality $1 - \mathrm{e}^{-x} \geq \frac{1}{2}\min(1,x)$, $x \geq 0$, we have

$$\left|1 - \mathrm{e}^{-2\kappa(s,\lambda_m)(\gamma(s)+R)}\right| \lesssim \max\left(1, \frac{T}{\bar{\sigma}L^2(1+RT^{-1})}\right) \lesssim \max\left(1, \frac{T}{\bar{\sigma}L^2}\right).$$

Combining the above bound and the bound of Proposition 3.9 into (39) yields

$$\|e_m^\sigma\|_{L^\infty(0,T;L^2(0,R))} \lesssim \mathrm{e}^{-4L^2\bar{\sigma}T^{-1}} \frac{T^{3/2}}{\bar{\sigma}L^2} \max\left(1, \frac{T}{\bar{\sigma}L^2}\right) \|\partial_t H_m(.,R)\|_{L^2(0,T)}.$$

Case $t \leq 2L$. Again, following the same argument as used in [BK21], it is sufficient to verify that $\|\hat{e}_m^\sigma(s)\|_{L^2(0,R)} \lesssim \mathrm{e}^{-2L\mathrm{Re}s}|s|^k \max(1, (\mathrm{Re}s)^\ell)$, for some $k, \ell \geq 0$. This is straightforward from (26):

$$\|\hat{e}_m^\sigma(s)\|_{L^2(0,R)} \leq \|c_m(s)\|_{L^2(0,R)}\mathrm{e}^{-2\mathrm{Re}\kappa(s,\lambda_m)\gamma(s)} \leq \|c_m(s)\|_{L^2(0,R)}\mathrm{e}^{-2\mathrm{Re}sL},$$

where the last bound follows from the definition of $\gamma(s) = L(1 + \frac{\bar{\sigma}}{s\zeta})$ and the technical Lemmas 3.7 and 3.8 (the latter implying in particular that $\mathrm{Re}(\kappa(s,\lambda_m)/s\zeta(s)) > 0$). Finally, the bound on $c_m(s)$ follows from (38). Because $\hat{H}_m$ itself is a Laplace transform of a function of $TD(H^1(\Omega_{ph}))$ (this is seen in particular from the same argument as the one used to derive Corollary 2.4), we have the desired bound in s on $|\hat{H}_m(s,R)|$ from Theorem 2.2. It remains to consider the term

$$|1 - \mathrm{e}^{-2\kappa(s,\lambda_m)(\gamma(s)+R)}| \gtrsim 1 - \mathrm{e}^{-2\mathrm{Re}(\kappa(s,\lambda_m)(\gamma(s)+R))} \gtrsim 1 - \mathrm{e}^{-2\mathrm{Re}sL} \gtrsim \min(1, \mathrm{Re}s),$$

where the last two bounds were obtained from the technical Lemmas 3.7 and 3.8 and the inequality $1 - \mathrm{e}^{-x} \geq 1/2\min(1,x)$ used before.

□

4. Extension of the results: examples and counterexamples for a waveguide with a heterogeneity

Since the techniques for the stability and convergence analysis from the previous section are largely based on the modal decomposition, it is natural to ask whether the respective results can be extended to waveguides with a perturbation (e.g. when ε and μ depend on the spatial variable away from the PMLs, when Ω_{ph} is no longer a rectangle, but a curved waveguide, or when there is an inclusion present in the domain). It is not difficult to verify that once the stability of the PML system is proven, the convergence can be shown by comparing the DtN operators, cf. the equation (24). E.g., when ε, μ depend on $\boldsymbol{x}$, one can start by proving the stability of the problem satisfied by the error of the PML. Provided the data g, this problem in the Laplace domain can be written as:

$$s^2\mu(s,\boldsymbol{x})\hat{e}^\sigma - \mathrm{div}\,\varepsilon^{-1}(s,\boldsymbol{x})\nabla\hat{e}^\sigma = 0 \quad \text{in } \Omega_{ph}, \tag{40}$$

$$\nabla\hat{e}^\sigma \cdot \boldsymbol{n} = 0 \text{ on } \partial\Omega_{ph} \setminus \Sigma, \qquad \nabla\hat{e}^\sigma \cdot \boldsymbol{n} = T^\sigma\hat{e}^\sigma + \hat{g}. \tag{41}$$

Next, one uses the same arguments as in the proof of Theorem 1.12 to show that in the case when $\hat{g} = (T^\sigma - T^\infty)\hat{H}$, the error is small thanks to the stability and the 'smallness' of the error between the two DtNs.

The main difficulty is to prove stability in this case. In this section we present two counter-examples, where we will prove that even if the heterogeneity is bounded away from the PMLs, the solution can exhibit instabilities in the time domain.

4.1. **Piecewise-constant dispersive media**

Let us consider wave propagation in media given by parameters of the following form:

$$\varepsilon(s,\boldsymbol{x}) = 1 + \frac{\omega_e^2(\boldsymbol{x})}{s^2}, \quad \omega_e(\boldsymbol{x}) = \text{const} > 0 \text{ for } x < R/2, \quad \omega_e(\boldsymbol{x}) = 0 \text{ for } x > R/2,$$
$$\mu(s,\boldsymbol{x}) = 1 + \frac{\omega_m^2(\boldsymbol{x})}{s^2}, \quad \omega_m(\boldsymbol{x}) = \text{const} > 0 \text{ for } x < R/2, \quad \omega_m(\boldsymbol{x}) = 0 \text{ for } x > R/2.$$

We consider the boundary-value problem written in the Laplace domain:

$$\begin{aligned} &s^2\mu\hat{H}^\sigma - \operatorname{div}\varepsilon^{-1}\nabla\hat{H}^\sigma = 0, \quad x > 0,\\ &\nabla\hat{H}^\sigma\cdot\boldsymbol{n} = 0 \text{ on } \mathbb{R}_*^+\times\Gamma,\\ &\nabla\hat{H}^\sigma\cdot\boldsymbol{n} = g \text{ on } \{0\}\times\Gamma. \end{aligned}$$

Let us remark that the time-domain counterpart of the above problem is stable, as can be shown by energy techniques.

Next, we apply a PML with the parameter ζ (because the media parameters are piecewise-constant rather than constant as before, we do not impose the conditions of Assumption 1.11 on ζ). We thus obtain

$$\begin{aligned} &s^2\mu\left(1+\frac{\sigma}{s\zeta}\right)\hat{H}^\sigma - \partial_x\varepsilon^{-1}\left(1+\frac{\sigma}{s\zeta}\right)^{-1}\partial_x\hat{H}^\sigma - \varepsilon^{-1}\left(1+\frac{\sigma}{s\zeta}\right)\partial_y^2\hat{H}^\sigma = 0, \quad x > 0,\\ &\nabla\hat{H}^\sigma\cdot\boldsymbol{n} = 0 \text{ on } (0,R+L)\times\Gamma\cup\{x=R+L\}, \quad \nabla\hat{H}^\sigma\cdot\boldsymbol{n} = g \text{ on } \{0\}\times\Gamma. \end{aligned} \tag{42}$$

Because the PML and the dispersive media are separated by a layer of vacuum, the so-called backward propagating waves which cause the instability of the PMLs do not reach the absorbing layer. Therefore one would expect that a 'good' (leading to a stable problem) choice of a PML would be the classical, Bérenger's PML with $\zeta = 1$.

On the other hand, arguing like in Section 2.2, one can formally reduce the above problem to the problem on the subdomain $(0,R)\times(0,\ell)$ with the boundary condition at $x = R$ written in the form $\varepsilon^{-1}\partial_x\hat{H}^\sigma = \widetilde{T}^\sigma\hat{H}^\sigma$, where $\widetilde{T}^\sigma$ is the DtN associated to the coupled vacuum-PML problem posed on $(R/2, R+L)\times(0,\ell)$. From this point of view it is less clear whether it is the PML change of variables with $\zeta = 1$ or ζ satisfying Assumption 1.11 that leads to a stable problem.

In this short section we will argue, semi-analytically and numerically, that neither of the choices leads to a stable formulation in this case.

For this it is sufficient to show that there exists $s_0 \in \mathbb{C}_*^+$, s.t. that the solution to (42) is not unique.

Thus, we will look for the solution to the problem (42) in the following form:

$$H^\sigma(s,x,y) = \begin{cases} \sum\limits_{m=0}^{\infty} g_m\left(a_m^+\mathrm{e}^{\kappa_d(s,\lambda_m)x} + a_m^-\mathrm{e}^{-\kappa_d(s,\lambda_m)x}\right)\phi_m(y), & x < \frac{R}{2},\\ \sum\limits_{m=0}^{\infty} g_m\left(b_m^+\mathrm{e}^{\kappa_0(s,\lambda_m)(x-R/2)} + b_m^-\mathrm{e}^{-\kappa_0(s,\lambda_m)(x-R/2)}\right)\phi_m(y), & \frac{R}{2} < x < R,\\ \sum\limits_{m=0}^{\infty} g_m\left(c_m^+\mathrm{e}^{\kappa_0(s,\lambda_m)(x_\sigma-R)} + c_m^-\mathrm{e}^{-\kappa_0(s,\lambda_m)(x_\sigma-R)}\right)\phi_m(y), & R < x < R+L, \end{cases}$$
$$x_\sigma(s) = x + \frac{1}{s\zeta}\int_R^x \sigma(x')dx',$$

where $\kappa_d(s,\lambda) = \sqrt{s^2\varepsilon(s)\mu(s)+\lambda^2}$ and $\kappa_0(s,\lambda) = \sqrt{s^2+\lambda^2}$.

The coefficients satisfy the following equation, which is obtained from the boundary conditions at $x=0$, the transmission conditions at $x=\frac{R}{2}$, the transmission conditions at $x=R$ and the boundary condition at $x=R+L$:

$$\begin{pmatrix} M_m^{(11)} & 0 & 0 \\ M_m^{(21)} & M_m^{(22)} & 0 \\ 0 & M_m^{(32)} & M_m^{(33)} \\ 0 & 0 & M_m^{(44)} \end{pmatrix} \begin{pmatrix} a_m^+ \\ a_m^- \\ b_m^+ \\ b_m^- \\ c_m^+ \\ c_m^- \end{pmatrix} = \begin{pmatrix} g_m \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

where the matrix M_m is given by

$$\begin{aligned}
M_m^{(11)} &= \begin{pmatrix} 1 & -1 \end{pmatrix}, \\
M_m^{(21)} &= \begin{pmatrix} \mathrm{e}^{\kappa_d(s,\lambda_m)\frac{R}{2}} & \mathrm{e}^{-\kappa_d(s,\lambda_m)\frac{R}{2}} \\ \frac{\kappa_d(s,\lambda_m)}{\varepsilon}\mathrm{e}^{\kappa_d(s,\lambda_m)\frac{R}{2}} & -\frac{\kappa_d(s,\lambda_m)}{\varepsilon}\mathrm{e}^{-\kappa_d(s,\lambda_m)\frac{R}{2}} \end{pmatrix}, \qquad M_m^{(22)} = \begin{pmatrix} -1 & -1 \\ -\kappa_0(s,\lambda_m) & \kappa_0(s,\lambda_m) \end{pmatrix}, \\
M_m^{(32)} &= \begin{pmatrix} \mathrm{e}^{\kappa_0(s,\lambda_m)\frac{R}{2}} & \mathrm{e}^{-\kappa_0(s,\lambda_m)\frac{R}{2}} \\ \kappa_0(s,\lambda_m)\mathrm{e}^{\kappa_0(s,\lambda_m)\frac{R}{2}} & -\kappa_0(s,\lambda_m)\mathrm{e}^{-\kappa_0(s,\lambda_m)\frac{R}{2}} \end{pmatrix}, \qquad M_m^{(33)} = \begin{pmatrix} -1 & -1 \\ -\kappa_0(s,\lambda_m) & \kappa_0(s,\lambda_m) \end{pmatrix}, \\
M_m^{(44)} &= \begin{pmatrix} \kappa_0(s,\lambda_m)\mathrm{e}^{\kappa_0(s,\lambda_m)\gamma(s)} & -\kappa_0(s,\lambda_m)\mathrm{e}^{-\kappa_0(s,\lambda_m)\gamma(s)} \end{pmatrix}.
\end{aligned}$$

The matrices $M_m^{(21)}$ and $M_m^{(33)}$ are invertible for all $s\in\mathbb{C}_*^+$ (their determinants are respectively $-2\kappa_d(s,\lambda_m)/\varepsilon$ and $-2\kappa_0(s,\lambda_m)$, which do not vanish for any $s\in\mathbb{C}_*^+$ by passivity of $\varepsilon(s)$ and Lemma 3.7). We thus express $\boldsymbol{a}_m = (a_m^+ \quad a_m^-)^T$ and $\boldsymbol{c}_m = (c_m^+ \quad c_m^-)^T$ via $\boldsymbol{b}_m = (b_m^+ \quad b_m^-)^T$. This allows us to reduce the problem to a linear equation on $\boldsymbol{b}_m$:

$$\begin{pmatrix} -\frac{\varepsilon}{2\kappa_d}\left(\mathrm{e}^{-\kappa_d R/2}\left(\frac{\kappa_d}{\varepsilon}+\kappa_0\right) - \mathrm{e}^{\kappa_d R/2}\left(\frac{\kappa_d}{\varepsilon}-\kappa_0\right)\right) & -\frac{\varepsilon}{2\kappa_d}\left(\mathrm{e}^{-\kappa_d R/2}\left(\frac{\kappa_d}{\varepsilon}-\kappa_0\right) - \mathrm{e}^{\kappa_d R/2}\left(\frac{\kappa_d}{\varepsilon}+\kappa_0\right)\right) \\ \mathrm{e}^{\kappa_0(\gamma_m+R/2)} & -\mathrm{e}^{-\kappa_0(\gamma_m+R/2)} \end{pmatrix} \boldsymbol{b}_m = \begin{pmatrix} g_m \\ 0 \end{pmatrix}.$$

The above matrix is not invertible if and only if, with $\xi = \frac{\kappa_d-\varepsilon\kappa_0}{\kappa_d+\varepsilon\kappa_0}$,

$$F(s) = \mathrm{e}^{-\kappa_0\gamma_m}\left(\mathrm{e}^{-(\kappa_d+\kappa_0)R/2} - \xi\mathrm{e}^{(\kappa_d-\kappa_0)R/2}\right) + \mathrm{e}^{\kappa_0\gamma_m}\left(\xi\mathrm{e}^{(\kappa_0-\kappa_d)R/2} - \mathrm{e}^{(\kappa_0+\kappa_d)R/2}\right) = 0. \tag{43}$$

Our goal is thus to show that there exists $s_0\in\mathbb{C}_*^+$, s.t. $F(s_0)=0$. Because F is analytic in $\mathbb{C}_*^+$, any such value of s_0 would be a root of F of a finite multiplicity.

Because estimating the sign of the real part of the roots of F seems fairly cumbersome, even in the asymptotic regime $\sigma\to 0$, we propose instead to look at the evolution of some of the roots, computed numerically, as σ increases.

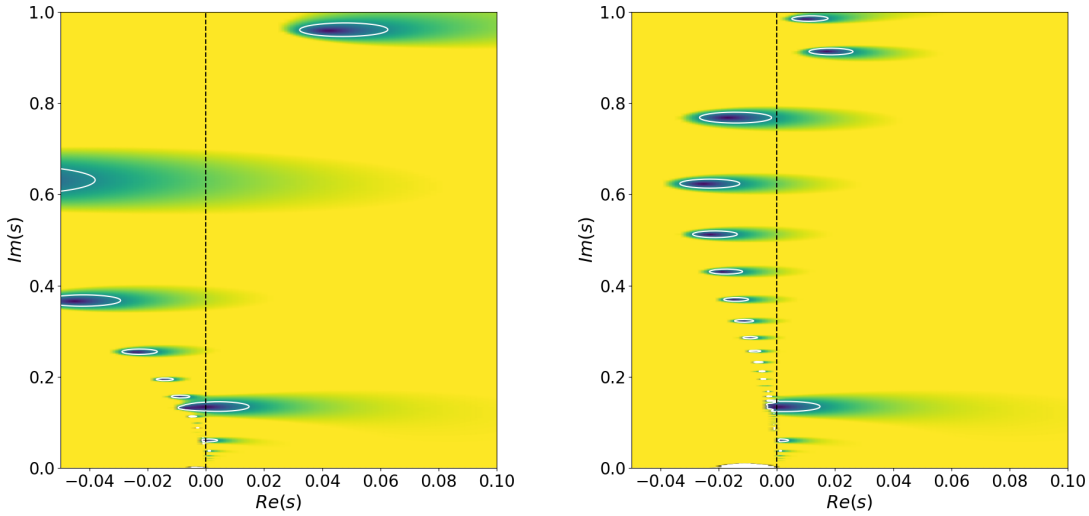
We proceed to show numerically that for a certain choice of parameters the function (43) has roots in $\mathbb{C}_*^+$. To this end we choose

$$\omega_e = 3, \qquad \omega_m = 0.2, \qquad R = 1, \qquad L = 0.5, \qquad \lambda_m = 1$$

and plot the values of $|F(s)|$ in the complex plane for different values of $\bar{\sigma}$. Note that due to the fact that F is holomorphic in the right half plane, we obtain by the minimum principle that a level set which is a closed curve in $\mathbb{C}_*^+$ must enclose a root of F. Figure 6 shows that such level sets exist for various choices of σ proving the existence of unstable solutions for the according PML truncation.

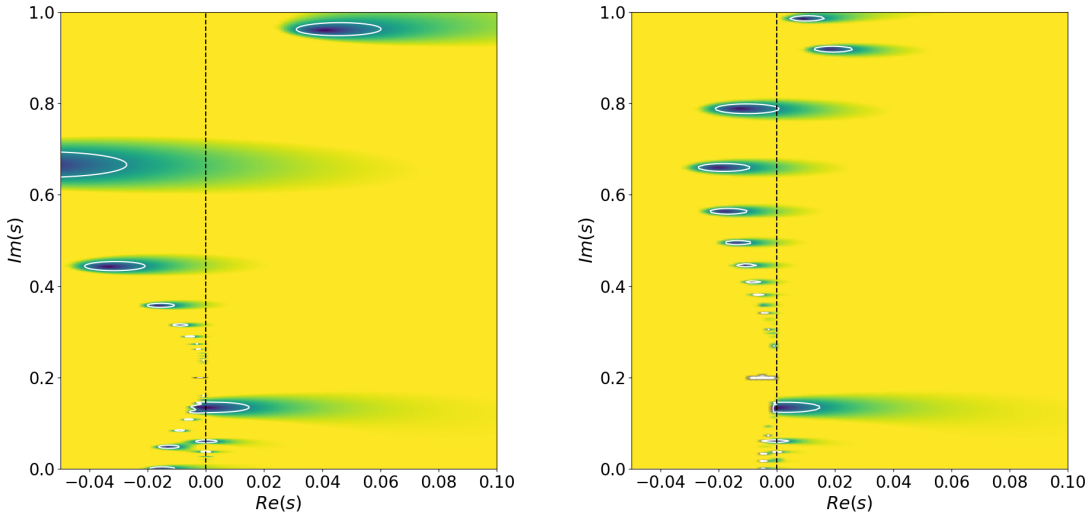
This is confirmed by computing the full two-dimensional time-domain simulation, where Figure 8a shows the instability in time.

Moreover Figure 7 shows that the choice of $\zeta=\mu$ which would be the 'correct' one for a homogeneous dispersive waveguide leads again to unstable solutions.



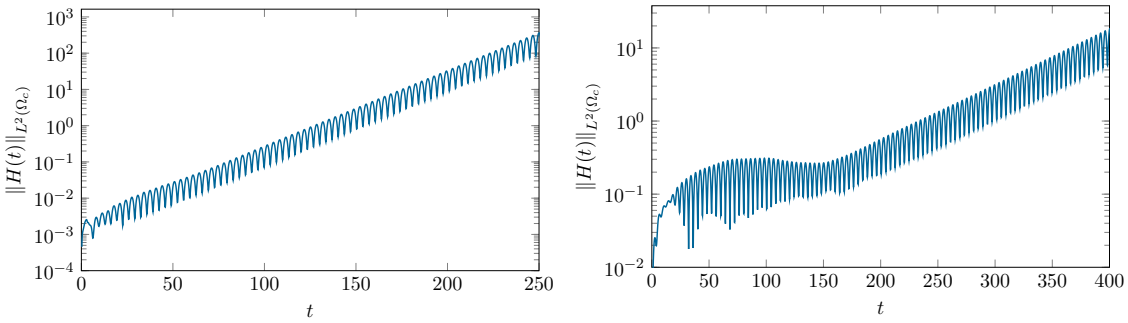
(A) $\bar{\sigma} = 5$ (B) $\bar{\sigma} = 15$

FIGURE 6. The values of $|F|$ (cf. (43), remember that here $\zeta = 1$) for various choices of $\bar{\sigma}$, where yellow indicates larger and blue smaller values. The white contours are the level sets corresponding to $|F| = 0.5$.



(A) $\bar{\sigma} = 5$ (B) $\bar{\sigma} = 15$

FIGURE 7. The values of the function corresponding to $|F|$ (cf. (43)) for $\zeta = \mu$ and for various choices of $\bar{\sigma}$, where yellow indicates larger and blue smaller values. The white contours are the level sets corresponding to $|F| = 0.5$.



(A) inhomogeneous, dispersive Drude materials with source frequency $\omega_0 = 1$ (for the remaining parameters see Sec. 4.1)

(B) inhomogeneous, non-dispersive materials with source frequency $\omega_0 = 0.8$ (for the remaining parameters see Sec. 4.2)

FIGURE 8. Exponential blow-up of the time-domain solution for the inhomogeneous waveguide problems with $\ell = \pi$, a time-harmonic source $f(t, x, y) = \sin(\omega_0 t) y \exp\left(-20(x^2 + (y - \pi/2)^2\right)$, a finite element mesh size $h = 0.1$, third order finite elements, a time step $\tau = 0.04$ and an explicit time-stepping scheme.

4.2. **Piecewise-constant non-dispersive media**

The results of the previous section may seem quite surprising, since the instability in the PMLs is produced even despite the fact that backward propagating waves do not reach the absorbing layer. This observation has incited us to ask whether the PMLs can be unstable when the heterogeneities do not support backward propagating waves at all. The goal of this section is to demonstrate semi-analytically and numerically that this may be indeed the case. For this we will relax the high-frequency requirement on μ and ε (i.e. we do not require that $\varepsilon(s)$, $\mu(s) \to 1$ as $s \to +\infty$), and will study the problem (42) with

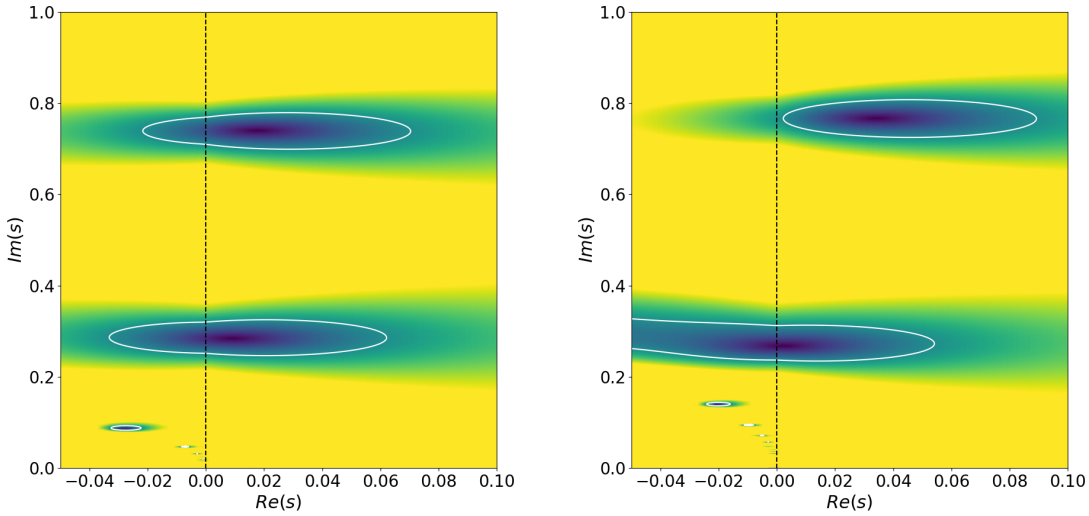
$$\varepsilon(s, \boldsymbol{x}) = \begin{cases} \varepsilon_H, & \text{if } x < R/2, \\ 1, & x > R/2, \end{cases} \qquad \mu(s, \boldsymbol{x}) = \begin{cases} \mu_H, & \text{if } x < R/2, \\ 1, & x > R/2, \end{cases} \quad \mu_H = \text{const} > 0, \quad \varepsilon_H = \text{const} > 0.$$

In this case (42) describes wave propagation in two media characterized by different wave velocities. Evidently, both decoupled problems are isotropic and non-dispersive; the stability of (42) in the time domain is straightforward from energy estimates. It is thus natural to apply the standard, Bérenger's PMLs to bound the computational domain, i.e., we choose $\zeta = 1$. Repeating the same arguments as in the previous section, we conclude that PML instabilities may occur if there exists the solution $s_0 \in \mathbb{C}_*^+$ to (43).

For this we choose the parameters as follows:

$$\varepsilon_H = \mu_H = 10, \quad R = 1, \quad L = 0.5, \quad \lambda_m = 1.$$

As in the non-dispersive case we observe that there exist roots (cf. Figure 9) of the resulting function F (suitably defined for the non dispersive case) in $\mathbb{C}_*^+$ for certain parameters. Again this shows the existence of unstable solutions (cf. Figure 8b).



(A) $\bar{\sigma} = 6$ (B) $\bar{\sigma} = 18$

FIGURE 9. The values of the function corresponding to $|F|$ (cf. (43)) for the non dispersive case (cf. Sec. 4.2) and various choices of $\bar{\sigma}$, where yellow indicates larger and blue smaller values. The white contours are the level sets corresponding to $|F| = 0.5$.

5. Numerical experiments

To underline our theoretical findings, we conduct numerical experiments studying the convergence of the truncation error of the PMLs. To this end we use an implementation of the PML equations (11) by high-order conforming finite elements on a triangular mesh using the software package NGSolve [Sch97, Sch14].

5.1. **Description of the experiments**

We consider a one-sided waveguide with physical domain $\Omega_{ph} = (0, R) \times (-1, 1)$ and PML domain $\Omega_\sigma = (R, R + L) \times (0, 1)$ for $R = 0.5$ and $L > 0$. In all the experiments we used finite elements of the polynomial order 7 (and higher orders for the reference solution) and a mesh size $h = 0.1$. We apply embedded Runge-Kutta 2(3) time-stepping methods to ensure that the time integration error is small.

For our experiments we choose Lorentz materials with a single pole of the form (cf. Figure 10)

$$\mu(s) = 1 + \frac{\mu_1^2}{s_{\mu,1}^2 + s^2}, \qquad \varepsilon(s) = 1 + \frac{\varepsilon_1^2}{s_{\varepsilon,1}^2 + s^2}$$

with $\mu_1, \varepsilon_1 = 10$ and $s_{\mu,1} = 6, s_{\varepsilon,1} = 8$.

Let us consider the dispersion relation of the plane waves $\mathrm{e}^{i(\omega t - \boldsymbol{k}\boldsymbol{x})}$ propagating in the free space media characterized by μ and ε, namely,

$$\mathcal{F}(\omega, \boldsymbol{k}) = \omega^2 \varepsilon(-i\omega)\mu(-i\omega) - |\boldsymbol{k}|^2 = 0. \tag{44}$$

It can be shown that the above implicit equation defines six branches of the solution $|\boldsymbol{k}| \mapsto \omega_j(|\boldsymbol{k}|)$, $j = 1, \ldots, 6$. Depending on the range of frequencies, the following types of wave propagation can be observed (cf. [CJK17] and [BJV18]):

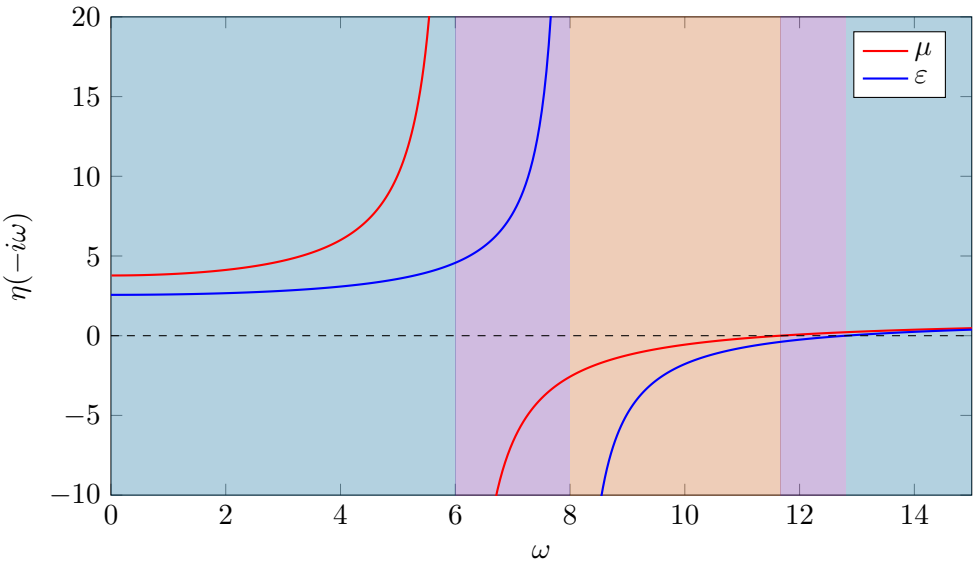


FIGURE 10. Materials used in the numerical experiments. The colors indicate the regions where we expect propagating, backward and evanescent waves.

- for ω s.t. $\mu(\omega) \geq 0$, $\varepsilon(\omega) \geq 0$ all the waves are forward propagating, more precisely, the group velocity $\boldsymbol{v}_g = \nabla_{\boldsymbol{k}}\omega(\boldsymbol{k})$ and the phase velocity $\boldsymbol{v}_{ph} = \frac{\omega}{|\boldsymbol{k}|}\boldsymbol{k}$ satisfy:

$$(\boldsymbol{v}_g \cdot \boldsymbol{e}_x)(\boldsymbol{v}_{ph} \cdot \boldsymbol{e}_x) \geq 0, \quad (\boldsymbol{v}_g \cdot \boldsymbol{e}_y)(\boldsymbol{v}_{ph} \cdot \boldsymbol{e}_y) \geq 0.$$

- for ω s.t. $\mu(\omega) < 0$, $\varepsilon(\omega) < 0$ all the waves are backward propagating, more precisely

$$(\boldsymbol{v}_g \cdot \boldsymbol{e}_x)(\boldsymbol{v}_{ph} \cdot \boldsymbol{e}_x) < 0, \quad (\boldsymbol{v}_g \cdot \boldsymbol{e}_y)(\boldsymbol{v}_{ph} \cdot \boldsymbol{e}_y) < 0.$$

- finally, for ω s.t. $\mu(\omega)\varepsilon(\omega) < 0$, the dispersion relation (44) has no real solutions, and thus the waves are evanescent.

This analysis originally done in the free-space dispersive medium can be extended to the case of the dispersive waveguide. As one can expect, cf. the seminal work [BFJ03] and a rigorous justification in [BJV18], the Bérenger's PML instabilities occur due to the presence of the backward propagating waves. The choice of the scaling function $\zeta = \mu$ stabilizes the perfectly matched layer.

The PML damping is chosen as $\sigma(x) \equiv \sigma_0$ which leads to the average damping $\bar{\sigma} = \sigma_0$.

Moreover, we introduce a time-dependent source by equipping the problem with the Neumann boundary values at $x = 0$ for H^σ

$$\nabla H^\sigma(t, 0, y) \cdot \boldsymbol{n} = \exp(-10y^2)\tau(t), \qquad \tau(t) = \frac{\sin(\omega_0 t)}{t} \exp(-\frac{1}{2}\ln^2(t)), \quad \omega_0 > 0.$$

To obtain the reference solution, we solve the same problem in the domain $(0, R + L_\infty) \times (0, \ell)$ with a homogeneous Neumann boundary condition for H^σ on the right end and L_∞ chosen so that for the considered time no reflected wave re-enters the interior domain. We measure the relative (with respect to the reference solution) error of H^σ in the $L^2(0, T; L^2(\Omega_{ph}))$-norm inside the interior domain only.

Remark 5.1. Due to the fact that the source has a fixed distance R to the start of the PML, we also might consider the whole waveguide as a PML of length $L + R$ with

$$\sigma(x) = \begin{cases} 0, & x \leq R, \\ \sigma_0, & x > R. \end{cases}$$

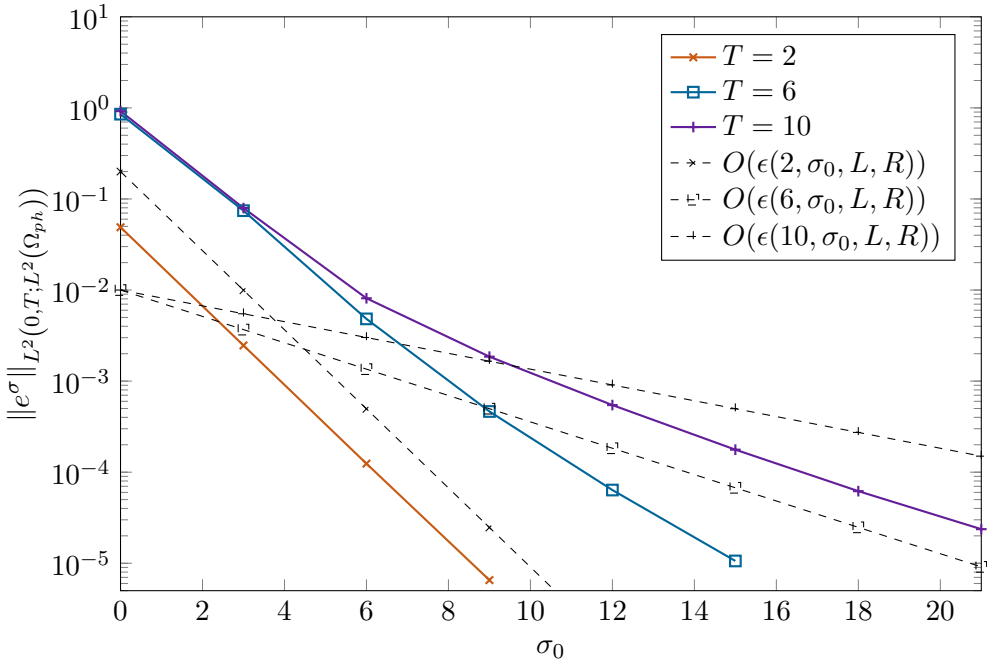


FIGURE 11. Experiment 1: Convergence of the PML-truncation for propagating waves with respect to the damping parameter. The discretization error is $\approx 10^{-6}$.

This leads to an average damping $\bar{\sigma} = \frac{L}{L+R}\sigma_0$ and thus to the exponential part of the error bound

$$\epsilon(T, \sigma_0, L, R) = \exp\left(-\frac{4L(L+R)\sigma_0}{T}\right).$$

5.2. **Experiments and the related discussion**

Experiment 1. In our first experiment we fix the PML length to $L = 0.5$, the frequency $\omega_0 = 20$, and vary the damping parameter σ_0, as well as the final time $T \in \{2, 6, 10\}$. Figure 11 shows the exponential convergence of the PML-truncation with respect to σ_0 compared to the theoretical estimates. We also observe that the asymptotic bound is sharp for $T = 2$.

Experiment 2: We fix $\sigma_0 = 4$, $\omega_0 = 20$, and vary the length of the PML L, as well as the final time $T \in \{3, 4, 5\}$. The results are shown in Figure 12. Again, we observe the exponential convergence in $L(R+L)$, whose rate is faster than the theoretical asymptotic rate.

Experiment 3: For the last experiment we vary the frequency $\omega_0 = 7, 10, 20$ of the source to obtain solutions in the three different regimes (cf. Figure 13) and study again the convergence in the PML damping σ_0. For comparison, we have added to the figure the PML error measured in the analogous configuration with a non-dispersive material. The observed exponential rate of convergence in this example depends very mildly on the source frequency, but exhibits quite significant dependence on the material. Quite surprisingly, the PMLs for the dispersive media show a better convergence rate than in the classical Bérenger's PMLs in the vacuum.

The above results show that the obtained theoretical error bounds predict the exponential convergence of the PML, however, the numerically observed convergence rates do not always coincide with the theoretical ones (cf. Figure 11). This may be related to the way we derived the error bounds: by taking the maximal value of the error indicator $\exp(A(s, \lambda_m, t))$ along the contour $s = s_0 + i\omega$, while, as we see in Figure 4, if the source contains only high frequencies in space and is e.g., time frequency band limited, it may happen that an actual PML error is much smaller than the one that was predicted by Theorem 1.12.

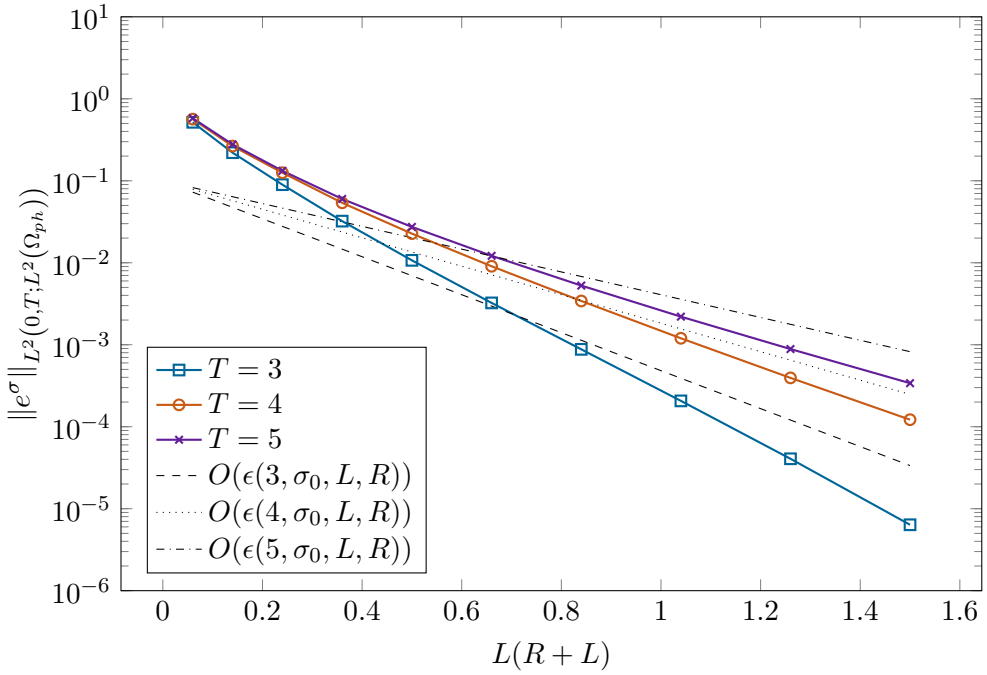


FIGURE 12. Experiment 2: Convergence of the PML-truncation for propagating waves with respect to the PML thickness. The discretization error is $\approx 10^{-6}$.

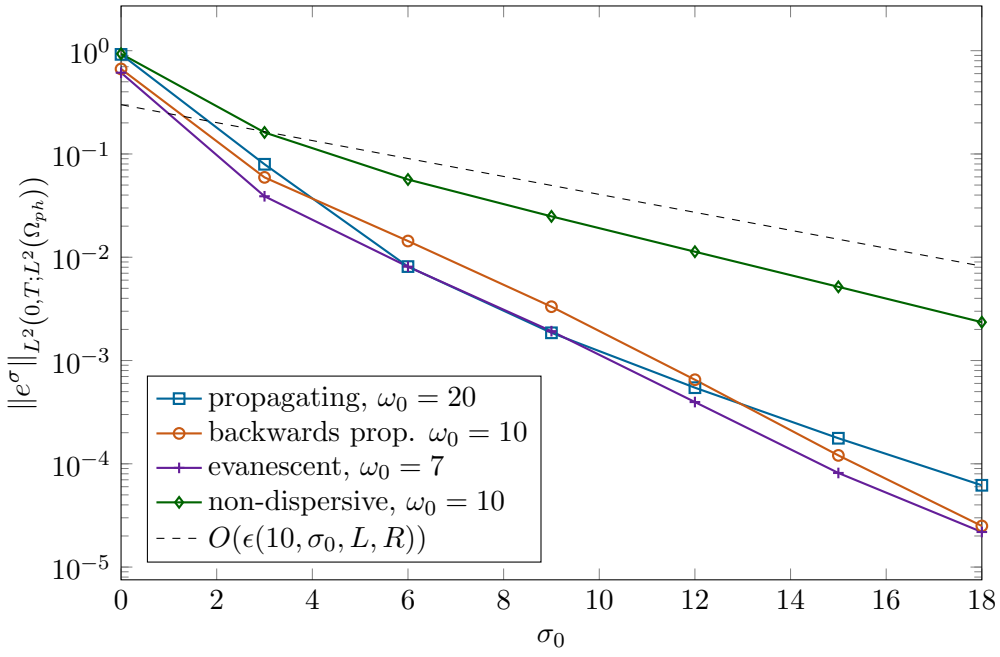


FIGURE 13. Experiment 3: Convergence of waves in different regimes. As a comparison the error of a non-dispersive waveguide is given.

6. Conclusions and open questions

In this article we have studied the convergence of generalized perfectly matched layers for the 2D electromagnetic wave propagation in homogeneous dispersive media. We have obtained the same exponential convergence bound as in the case of Bérenger's PMLs for non-dispersive media. Similar arguments can of course be applied to the 2D/3D acoustic wave propagation in dispersive waveguides.

Our arguments can be extended to the case of waveguides with localized heterogeneities, as soon as stability is proven. However, as we show using semi-analytic arguments and numerical experiments, even in the case of localized heterogeneity in non-dispersive waveguides, the PMLs (Bérenger's PMLs) may exhibit instabilities(!). It is thus an open question how to construct stable PMLs for these cases.

Another interesting open question is investigating the stability and convergence of PMLs for 3D Maxwell waveguides, both in the dispersive and the non-dispersive case.

References

[ADGK03] Sergey Asvadurov, Vladimir Druskin, Murthy N. Guddati, and Leonid Knizhnerman. On optimal finite-difference approximation of PML. SIAM J. Numer. Anal., 41(1):287–305, 2003.

[AG97] S. Abarbanel and D. Gottlieb. A mathematical analysis of the PML method. J. Comput. Phys, 134(2):357–363, 1997.

[AGH99] S. Abarbanel, D. Gottlieb, and J. S. Hesthaven. Well-posed perfectly matched layers for advective acoustics. J. Comput. Phys., 154(2):266–283, 1999.

[AGH02] S. Abarbanel, D. Gottlieb, and J.S. Hesthaven. Long Time Behavior of the Perfectly Matched Layer Equations in Computational Elactromagnetics. J. Scientific Comp., 17(1-4):405–422, 2002.

[AHK06] Daniel Appelö, Thomas Hagstrom, and Gunilla Kreiss. Perfectly matched layers for hyperbolic systems: general formulation, well-posedness, and stability. SIAM Journal on Applied Mathematics, 67(1):1–23, 2006.

[BBDCZ10] A. S. Bonnet-Ben Dhia, P. Ciarlet, Jr., and C. M. Zwölf. Time harmonic wave diffraction problems in materials with sign-shifting coefficients. J. Comput. Appl. Math., 234(6):1912–1919, 2010.

[BDT21] Nacéra Baara, Julien Diaz, and Mounir Tlemcani. Time domain analysis and localization of a non-local PML for dispersive wave equations. J. Comput. Phys., 445:Paper No. 110638, 18, 2021.

[Ber94] Jean-Pierre Berenger. A perfectly matched layer for the absorption of electromagnetic waves. J. Comput. Phys., 114(2):185–200, 1994.

[Ber96] Jean-Pierre Berenger. Three-dimensional perfectly matched layer for the absorption of electromagnetic waves. J. Comput. Phys., 127(2):363–379, 1996.

[Ber97] J.P. Berenger. Improved PML for the FDTD Solution of Wave-Structure Interaction Problems. IEEE transactions on Antennas and Propagation., 45(3):466–473, march 1997.

[BFJ03] E. Bécache, S. Fauqueux, and P. Joly. Stability of perfectly matched layers, group velocities and anisotropic waves. J. Comput. Phys., 188(2):399–433, 2003.

[BGIK19] Daniel H. Baffet, Marcus J. Grote, Sébastien Imperiale, and Maryna Kachanovska. Energy decay and stability of a perfectly matched layer for the wave equation. J. Sci. Comput., 81(3):2237–2270, 2019.

[BJ02] Eliane Bécache and Patrick Joly. On the analysis of Bérenger's perfectly matched layers for Maxwell's equations. M2AN Math. Model. Numer. Anal., 36(1):87–119, 2002.

[BJK17] Eliane Bécache, Patrick Joly, and Maryna Kachanovska. Stable perfectly matched layers for a cold plasma in a strong background magnetic field. J. Comput. Phys., 341:76–101, 2017.

[BJV18] Éliane Bécache, Patrick Joly, and Valentin Vinoles. On the analysis of perfectly matched layers for a class of dispersive media and application to negative index metamaterials. Math. Comp., 87(314):2775–2810, 2018.

[BK17] Eliane Bécache and Maryna Kachanovska. Stable perfectly matched layers for a class of anisotropic dispersive models. Part I: Necessary and sufficient conditions of stability. ESAIM Math. Model. Numer. Anal., 51(6):2399–2434, 2017.

[BK21] Eliane Bécache and Maryna Kachanovska. Stability and Convergence Analysis of Time-domain Perfectly Matched Layers for The Wave Equation in Waveguides. SIAM Journal on Numerical Analysis, 2021.

[BPG04] Eliane Bécache, Peter Petropoulos, and Stephen Gedney. On the long-time behavior of unsplit Perfectly Matched Layers. IEEE Transactions on Antennas and Propagation, 52(5):–, 2004.

[CC41] Kenneth S. Cole and Robert H. Cole. Dispersion and absorption in dielectrics i. alternating current characteristics. The Journal of Chemical Physics, 9(4):341–351, 1941.

[Che19] Albert Chern. A reflectionless discrete perfectly matched layer. J. Comput. Phys., 381:91–109, 2019.

[CJK17] Maxence Cassier, Patrick Joly, and Maryna Kachanovska. Mathematical models for dispersive electromagnetic waves: an overview. Comput. Math. Appl., 74(11):2792–2830, 2017.

[CM98] F. Collino and P. B. Monk. Optimizing the perfectly matched layer. volume 164, pages 157–171. 1998. Exterior problems of wave propagation (Boulder, CO, 1997; San Francisco, CA, 1997).

[Col97] F. Collino. Perfectly Matched Absorbing Layers for the Paraxial Equations . J. of Comp. Phys., 131(1):164–180, 1997.

[CSL10] Tie Jun Cui, David R Smith, and Ruopeng Liu. Metamaterials: theory, design, and applications. Springer, 2010.

[CT01] F. Collino and C. Tsogka. Application of the pml absorbing layer model to the linear elastodynamic problem in anisotropic heteregeneous media. Geophysics, 66(1):294–307, 2001.

[Cum04] Steven A Cummer. Perfectly matched layer behavior in negative refractive index materials. Antennas and Wireless Propagation Letters, IEEE, 3(1):172–175, 2004.

[DI13] E. Demaldent and S. Imperiale. Perfectly matched transmission problem with absorbing layers: application to anisotropic acoustics in convex polygonal domains. Internat. J. Numer. Methods Engrg., 96(11):689–711, 2013.

[DJ06a] J. Diaz and P. Joly. A time domain analysis of PML models in acoustics. Comput. Methods Appl. Mech. Engrg., 195(29-32):3820–3853, 2006.

[DJ06b] J. Diaz and P. Joly. A time domain analysis of PML models in acoustics. Comput. Methods Appl. Mech. Engrg., 195(29-32):3820–3853, 2006.

[DK12] Kenneth Duru and Gunilla Kreiss. On the accuracy and stability of the perfectly matched layer in transient waveguides. J. Sci. Comput., 53(3):642–671, 2012.

[Ged96] Stephen D Gedney. An anisotropic perfectly matched layer-absorbing medium for the truncation of fdtd lattices. IEEE transactions on Antennas and Propagation, 44(12):1630–1639, 1996.

[Hes98] J. S. Hesthaven. On the Analysis and Construction of Perfectly Matched Layers for the Linearized Euler Equations. J. Comp. Phys., 142:129–147, 1998.
[HJL18] Yunqing Huang, Hongen Jia, and Jichun Li. Analysis and application of an equivalent Berenger's PML model. J. Comput. Appl. Math., 333:157–169, 2018.
[HN67] S. Havriliak and S. Negami. A complex plane representation of dielectric and mechanical relaxation processes in some polymers. Polymer, 8:161–210, 1967.
[HPBR11] L. Halpern, S. Petit-Bergez, and J. Rauch. The analysis of matched layers. Confluentes Math., 3(2):159–236, 2011.
[HSB96] F. Hastings, J.B. Schneider, and S. L. Broschat. Application of the perfectly matched layer (PML) absorbing boundary condition to elastic wave propagation . J. Acoust. Soc. Am., 100:3061– 3069, 1996.
[Hu96] F. Q. Hu. On absorbing boundary conditions for linearized euler equations by a perfectly matched layer. J. Comp. Phys., 129:201–219, 1996.
[Hu01] Fang Q. Hu. A stable, perfectly matched layer for linearized Euler equations in unsplit physical variables. J. Comput. Phys., 173(2):455–480, 2001.
[MDG14] A. Modave, E. Delhez, and C. Geuzaine. Optimizing perfectly matched layers in discrete contexts. Internat. J. Numer. Methods Engrg., 99(6):410–437, 2014.
[Nat06] Frédéric Nataf. A new approach to perfectly matched layers for the linearized Euler system. J. Comput. Phys., 214(2):757–772, 2006.
[Ort90] James M. Ortega. Numerical analysis, volume 3 of Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second edition, 1990. A second course.
[Pet00] P. G. Petropoulos. Reflectionless sponge layers as absorbing boundary condition for the numerical solution of Maxwell's equations in rectangular, cylindrical, and spherical coordinates. SIAM J. Appl. Math., 60(3):1037–1058, 2000.
[PZC98] P. G. Petropoulos, L. Zhao, and A. C. Cangellaris. A reflectionless sponge layer absorbing boundary condition for the solution of Maxwell's equations with high-order staggered finite difference schemes. J. Comput. Phys., 139(1):184–208, 1998.
[Say16] Francisco-Javier Sayas. Retarded potentials and time domain boundary integral equations, volume 50 of Springer Series in Computational Mathematics. Springer, [Cham], 2016. A road map.
[Sch97] Joachim Schöberl. Netgen - an advancing front 2d/3d-mesh generator based on abstract rules. Comput.Visual.Sci, 1:41–52, 1997.
[Sch14] Joachim Schöberl. C++11 implementation of finite elements in ngsolve. Preprint 30/2014, Institute for Analysis and Scientific Computing, TU Wien, 2014.
[SPW04] DR Smith, JB Pendry, and MCK Wiltshire. Metamaterials and negative refractive index. Science, 305(5685):788–792, 2004.
[TAC98] C. K. W. Tam, L. Auriault, and F. Cambuli. Perfectly matched layer as an absorbing boundary condition for the linearized euler equations in open and ducted domains. J. Comput. Phys., 144(1):213–234, 1998.
[TC98] F. L. Teixeira and W. C. Chew. On Causality and Dynamic Stability of Perfectly Matched Layers for FDTD Simulations. Micro. Opt. Tech. Lett., 17:231–236, 1998.
[TY98] E. Turkel and A. Yefet. Absorbing PML boundary layers for wave-like equations. Appl. Numer. Math., 27(4):533–557, 1998.
[Vin16] Valentin Vinoles. Problèmes d'interface en présence de métamatériaux: modélisation, analyse et simulations. PhD thesis, Université Paris-Saclay (ComUE), 2016.
[WHNr78] G. Wanner, E. Hairer, and S. P. Nø rsett. Order stars and stability theorems. BIT, 18(4):475–489, 1978.
[ZC96] L. Zhao and A.C. Cangellaris. GT PML: Generalize Theory of Perfectly Matched Layers and Its Application to the Reflectionless Truncation of Finite-Difference Time-Domain Grids. IEEE Trans. Microwave Theory Tech., 44:2555–2563, 1996.

Appendix A. Proof of Lemma 2.3

We will use the following bound from Lemma 2.4 of [BK17]: for any admissible η, it holds that

$$|\eta(s)| \lesssim |s| \max(1, (\mathrm{Re} s)^{-3}), \quad s \in \mathbb{C}_*^+. \tag{45}$$

The proof below is nothing less but a generalization of the approach suggested in [BK21]. We start by considering the PML sesquilinear form written using the modal decomposition

$$a_s(u,v) = \int_0^{R+L} \sum_{m=0}^{\infty} \left(\left(s^2\varepsilon\mu + \lambda_m^2\right) \left(1 + \frac{\sigma}{s\zeta}\right) u_m \overline{v}_m + \left(1 + \frac{\sigma}{s\zeta}\right)^{-1} \partial_x u_m \, \partial_x \overline{v}_m \right) dx.$$

Let us introduce $\widetilde{\mu} := \frac{\varepsilon\mu}{\zeta}$, which, by Assumption 1.11, is admissible, cf. Definition 1.1. We then have

$$a_s(u,v) = \int_0^{R+L} \sum_{m=0}^{\infty} \left(\left(s^2\widetilde{\mu}\zeta + \lambda_m^2\right)\left(1+\frac{\sigma}{s\zeta}\right) u_m \overline{v}_m + \left(1+\frac{\sigma}{s\zeta}\right)^{-1} \partial_x u_m \, \partial_x \overline{v}_m \right) dx.$$

First of all, remark that for $s \in \mathbb{C}^+$, the form a_s is continuous, because, on one hand, $\zeta, \widetilde{\mu}$ are analytic in $\mathbb{C}^+$, and, on the other hand $(s\zeta+\sigma)$ is bounded for all $\mathrm{Re}s > 0$, $\sigma \geq 0$, since $\mathrm{Re}(s\zeta+\sigma) \geq \mathrm{Re}(s\zeta) > 0$, by Assumption 1.11.

Let us now define $\tilde{s}_i := \mathrm{Im}(s\widetilde{\mu})\mathrm{Im}(s\zeta)$. To prove the inf-sup condition, we consider two cases.

(1) $\tilde{s}_i \leq 0$. It is then straightforward to verify that $\mathrm{Re}\, a_s(u,u) \geq c_s \|u\|^2_{H^1(\Omega_c)}$. Indeed,

$$\begin{aligned}
\mathrm{Re}\, a_s(u,u) &= \int_0^{R+L} \sum_{m=0}^{\infty} \left(\mathrm{Re}(s\zeta)\mathrm{Re}(s\widetilde{\mu}) - \widetilde{s}_i + \lambda_m^2 + \sigma(x)\mathrm{Re}\left(s\widetilde{\mu}\right) + \sigma(x)\lambda_m^2 \mathrm{Re}(s\zeta)^{-1}\right) |u_m(x)|^2 dx \\
&\qquad + \int_0^{R+L} \sum_{m=0}^{\infty} \frac{|s\zeta|^2 + \sigma(x)\mathrm{Re}(s\zeta)}{|s\zeta+\sigma(x)|^2} |\partial_x u_m(x)|^2 dx \\
&\geq \int_0^{R+L} \sum_{m=0}^{\infty} ((\mathrm{Re}s)^2 + \lambda_m^2)|u_m(x)|^2 dx + \min(1, (\mathrm{Re}s)^2 \|\sigma\|_\infty^{-2}) \int_0^{R+L} \sum_{m=0}^{\infty} |u_m(x)|^2 dx,
\end{aligned}$$

where the result follows by recalling the positivity result of (3), $\widetilde{s}_i \leq 0$ and the following bound:

$$\mathrm{Re}\frac{|s\zeta|^2 + \sigma\mathrm{Re}(s\zeta)}{|s\zeta+\sigma|^2} \gtrsim \frac{|s\zeta|^2}{|s\zeta|^2 + \|\sigma\|_\infty^2} \gtrsim \frac{|\mathrm{Re}(s\zeta)|^2}{|\mathrm{Re}(s\zeta)|^2 + \|\sigma\|_\infty^2}.$$

The latter bound is obtained by remarking that the function $x \mapsto \frac{x}{x+a}$ grows in $x > 0$. For the same reason we can use (3) to bound $\mathrm{Re}(s\zeta)$ from below by $\mathrm{Re}s$:

$$\mathrm{Re}\frac{|s\zeta|^2 + \sigma\mathrm{Re}(s\zeta)}{|s\zeta+\sigma|^2} \gtrsim \frac{(\mathrm{Re}s)^2}{(\mathrm{Re}s)^2 + \|\sigma\|_\infty^2} \gtrsim \frac{(\mathrm{Re}s)^2}{\max((\mathrm{Re}s)^2, \|\sigma\|_\infty^2)} \gtrsim \min(1, \left(\frac{\mathrm{Re}s}{\|\sigma\|_\infty}\right)^2). \tag{46}$$

(2) $\tilde{s}_i > 0$. We will prove an inf-sup condition in terms of the T-coercivity, cf. [BBDCZ10]. We define the operator $T_s : H^1(\Omega_c) \to H^1(\Omega_c)$ by

$$T_s v = \sum_{\lambda_m^2 < \widetilde{s}_i} s\zeta(s) v_m \phi_m + \sum_{\lambda_m^2 \geq \widetilde{s}_i} v_m \phi_m.$$

Evidently, T_s is invertible with a bounded inverse. We then have

$$\begin{aligned}
\mathrm{Re}a(u, T_s u) = &\underbrace{\int_0^{R+L} \sum_{\lambda_m^2 < \widetilde{s}_i} \left(|s\zeta|^2 \mathrm{Re}(s\widetilde{\mu}) + \lambda_m^2 \mathrm{Re}(\overline{s\zeta}) + \frac{\sigma(x)}{|s\zeta|^2}\left(|s\zeta|^2 \mathrm{Re}(s\widetilde{\mu}\overline{s\zeta}) + \lambda_m^2 \mathrm{Re}(\overline{s\zeta})^2\right)\right) |u_m|^2}_{I_1} \\
&+ \underbrace{\int_0^{R+L} \sum_{\lambda_m^2 \geq \widetilde{s}_i} \left(\mathrm{Re}(s\zeta)\mathrm{Re}(s\widetilde{\mu}) - \widetilde{s}_i + \lambda_m^2 + \sigma(x)\left(\mathrm{Re}(s\widetilde{\mu}) + \lambda_m^2 \mathrm{Re}(s\zeta)^{-1}\right)\right) |u_m|^2}_{I_2} \\
&+ \int_0^{R+L} \sum_{\lambda_m^2 < \widetilde{s}_i} \mathrm{Re}\left(\frac{|s\zeta|^2}{s\zeta+\sigma}\right) |\partial_x u_m|^2 + \int_0^{R+L} \sum_{\lambda_m^2 \geq \widetilde{s}_i} \mathrm{Re}\left(\frac{|s\zeta|^2 + \sigma s\zeta}{|s\zeta+\sigma|^2}\right) |\partial_x u_m|^2
\end{aligned}$$

The latter term in the above is bounded from below using (46). The previous term can be bounded using (3) and (46)

$$\mathrm{Re}\left(\frac{|s\zeta|^2}{s\zeta+\sigma}\right) \gtrsim \frac{|s\zeta|^2\mathrm{Re}(s\zeta)}{|s\zeta+\sigma|^2} \gtrsim \mathrm{Re}s \min\left(1, \left(\frac{\mathrm{Re}s}{\|\sigma\|_\infty}\right)^2\right).$$

It remains to study the signs of I_1 and I_2. In I_1, the only term that is not a priori sign-definite (cf. Assumption 1.11) is developed below:

$$|s\zeta|^2\mathrm{Re}(s\widetilde{\mu}\overline{s\zeta}) + \lambda_m^2\mathrm{Re}(\overline{s\zeta})^2 \geq |s\zeta|^2\left(\mathrm{Re}(s\widetilde{\mu})\mathrm{Re}(s\zeta) + \widetilde{s}_i - \lambda_m^2\right) \geq |s\zeta|^2\mathrm{Re}(s\widetilde{\mu})\mathrm{Re}(s\zeta), \text{ because } \lambda_m^2 < \widetilde{s}_i \text{ in } I_1.$$

This yields the lower bound

$$I_1 \geq (\mathrm{Re}(s\zeta))^2\mathrm{Re}(s\widetilde{\mu})\|u\|^2 + \sum_{\lambda_m^2<\widetilde{s}_i} \mathrm{Re}(s\zeta)\lambda_m^2|u_m|^2 \geq (\mathrm{Re}s)^3\|u\|^2 + \mathrm{Re}s \sum_{\lambda_m^2<\widetilde{s}_i} \lambda_m^2|u_m|^2,$$

where the last bound follows from (3).

Let us now consider the term I_2. The term I_2 is clearly bounded from below by $\mathrm{Re}(s\zeta)\mathrm{Re}(s\widetilde{\mu})\|u\|^2 \gtrsim (\mathrm{Re}s)^2\|u\|^2$, since $\lambda_m^2 \geq \widetilde{s}_i$, and it remains to argue that $I_2 \gtrsim \sum_{\lambda_m^2\geq\widetilde{s}_i} \lambda_m^2|u_m|^2$.

For any $0 < \varepsilon < 1$ (to be chosen), it holds that

$$\begin{aligned} I_2 &\geq \int_0^{R+L} \sum_{\lambda_m^2(1-\varepsilon)\geq\widetilde{s}_i} ((\mathrm{Re}s)^2|u_m|^2 + (\lambda_m^2 - \widetilde{s}_i)|u_m|^2) + \int_0^{R+L} \sum_{(1-\varepsilon)\widetilde{s}_i\leq\lambda_m^2(1-\varepsilon)\leq\widetilde{s}_i} (\mathrm{Re}s)^2|u_m|^2 \\ &\geq \int_0^{R+L} \sum_{\lambda_m^2(1-\varepsilon)\geq\widetilde{s}_i} ((\mathrm{Re}s)^2|u_m|^2 + \varepsilon\lambda_m^2|u_m|^2) + \int_0^{R+L} \sum_{(1-\varepsilon)\widetilde{s}_i\leq\lambda_m^2(1-\varepsilon)\leq\widetilde{s}_i} \left(\frac{(\mathrm{Re}s)^2}{2}|u_m|^2 + \frac{1-\varepsilon}{2\widetilde{s}_i}\lambda_m^2(\mathrm{Re}s)^2|u_m|^2\right). \end{aligned}$$

Let us remark that $\widetilde{s}_i \leq |s|^4\max(1,(\mathrm{Re}s)^{-6})$, by (45). We choose e.g. $\varepsilon = 1/2$, so that we finally obtain the following bound:

$$\begin{aligned} I_2 &\gtrsim (\mathrm{Re}s)^2 \sum_{\lambda_m^2\geq\widetilde{s}_i} \|u_m\|^2_{L^2(0,R+L)} + \min\left(1, \frac{1}{|s|^4}\right)\min(1,(\mathrm{Re}s)^6) \sum_{\lambda_m^2\geq\widetilde{s}_i} \lambda_m^2|u_m^2| \\ &\gtrsim (\mathrm{Re}s)^2 \sum_{\lambda_m^2\geq\widetilde{s}_i} \|u_m\|^2_{L^2(0,R+L)} + |s|^{-4}\min(1,(\mathrm{Re}s)^{10}) \sum_{\lambda_m^2\geq\widetilde{s}_i} \lambda_m^2|u_m^2|, \end{aligned}$$

where the last bound follows from $\min(1,|s|^{-4}) \geq |s|^{-4}\min(|s|^4,1) \geq |s|^{-4}\min((\mathrm{Re}s)^4,1)$.

Finally, gathering all the bounds in the statement of the lemma, we notice that, for $\widetilde{s}_i \leq 0$,

$$|a_s(u,u)| \gtrsim (\mathrm{Re}s)^2\|u\|^2 + \|\partial_y u\|^2 + \|\sigma\|_\infty^{-2}\min(1,(\mathrm{Re}s)^2)\|\partial_x u\|^2 \geq C_\sigma \min(1,(\mathrm{Re}s)^2)\|u\|^2_{H^1(\Omega_c)}.$$

For $\widetilde{s}_i > 0$,

$$|a(u,T_s u)| \geq c_\sigma\left(\min(1,(\mathrm{Re}s)^3)\|\partial_x u\|^2 + (\mathrm{Re}s)^2\min(1,\mathrm{Re}s)\|u\|^2 + \min(\mathrm{Re}s, |s|^{-4}\min(1,(\mathrm{Re}s)^{10}))\|\partial_y u\|^2\right).$$

Because $\min(\mathrm{Re}s, |s|^{-4}\min(1,(\mathrm{Re}s)^{10})) \geq |s|^{-4}\min(1,(\mathrm{Re}s)^{10})$, we can conclude that

$$|a(u,T_s u)| \gtrsim c_\sigma|s|^{-4}\min(1,(\mathrm{Re}s)^{10})\|u\|^2_{H^1(\Omega_c)}.$$

Finally, the operator T_s is bounded, and we have that, by (45),

$$\|T_s\| \lesssim \max(1,|s|^2\max(1,(\mathrm{Re}s)^{-4})).$$

This implies the desired bound in the statement of the lemma.

Appendix B. Level sets of harmonic functions

For an open set $D \subset \mathbb{C}$, let $\phi : D \to \mathbb{C}$ be holomorphic. Then $\phi_R := \mathrm{Re}(\phi)$ is a harmonic function. Harmonic functions have the following well-known properties.

Proposition B.1. *Let $\phi_R : D \to \mathbb{R}$ be the real part of a holomorphic function ϕ. Then ϕ_R fulfills the min-max principle: for every simple closed curve $\mathcal{C} \subset D$ with $\mathcal{C} = \partial M$ for an open set $M \subset D$ we have*

$$\inf_{s\in M} \phi_R(s) = \min_{s\in \overline{M}} \phi_R(s) = \min_{s\in \mathcal{C}} \phi_R(s), \qquad \sup_{s\in M} \phi_R(s) = \max_{s\in \overline{M}} \phi_R(s) = \max_{s\in \mathcal{C}} \phi_R(s).$$

Proof of Proposition 3.3. (1) Follows from the implicit function theorem.

(2) Suppose that a level set contains a closed curve. Then by the min-max-principle it would follow that the function ϕ_R is constant in the interior of the said curve. From the Cauchy-Riemann equations the same follows for ϕ and thus ϕ would have to be constant also in the whole domain.

(3) A level set cannot be compact since it would be a closed curve in this case.

(4) The proof is similar to the reasoning of Lemma 3 in [WHNr78]. Let $z_0 \in D$ be a stationary point of ϕ, i.e., $\phi'(z_0) = 0$. Then, since ϕ is not constant there exists an index $m \in \mathbb{N}, m \geq 2$, such that

$$\phi(z) = \phi(z_0) + \phi_m (z - z_0)^m + O((z - z_0)^{m+1}),$$

for some $\phi_m \neq 0$ and z sufficiently close to z_0. Thus

$$\phi_R(z_0 + \varepsilon \exp(i\theta)) - \phi_R(z_0) = |\phi_m| \varepsilon^m \cos(m\theta + \arg(\phi_m)) + O(\varepsilon^{m+1}),$$

for $\varepsilon > 0$ small enough and $\theta \in [0, 2\pi)$. We now may choose $\varepsilon_0 > 0$ small enough such that the function $\theta \mapsto \phi_R(z_0 + \varepsilon \exp(i\theta)) - \phi_R(z_0)$ has $2m$ different roots in $[0, 2\pi)$. This shows that $2m$ branches of the level set meet in the point s_0. Because for sufficiently small ε in $B_\varepsilon(s_0)$ they do not intersect but in s_0, the statement follows.

□

Appendix C. Proof of Proposition 3.4

Let us recall the result to prove.

Proposition C.1. *For all $\bar\sigma > 0$, $L > 0$, $t > 2L$, $\lambda > 0$, the function $s \mapsto (st - 2\gamma(s)\kappa(s, \lambda))'$ has precisely two complex conjugate roots with a positive real part.*

Let us explain our strategy to prove the above proposition. First,

$$D(s) = (st - 2\gamma(s)\kappa(s,\lambda))' = t - \frac{2sL}{\sqrt{s^2+\lambda^2}}\left(1 + \frac{\bar\sigma}{s}\right) + \frac{2\bar\sigma L}{s^2}\sqrt{s^2+\lambda^2} = t - \frac{2Ls}{\sqrt{s^2+\lambda^2}} + \frac{2L\bar\sigma\lambda^2}{s^2\sqrt{s^2+\lambda^2}}.$$

Because the number of roots in $\mathbb{C}^+_*$ remains constant under multiplication by $\sqrt{s^2+\lambda^2}s^2$, instead we can consider

$$p(s) = ts^2\sqrt{s^2+\lambda^2} - 2Ls^3 + 2L\bar\sigma\lambda^2.$$

We proceed as follows. By a suitable rescaling $s = \lambda z$, we have that

$$p(s) = \lambda^3 \widetilde{p}(z), \quad \widetilde{p}(z) = tz^2\sqrt{z^2+1} - 2Lz^3 + 2L\frac{\bar\sigma}{\lambda}.$$

Let $L > 0$ be fixed, and let us consider the family of the analytic in $\mathbb{C} \setminus i\mathbb{R}$ and continuous in $\mathbb{C} \setminus \{z : z = ix, x \in \mathbb{R}, |x| \geq 1\}$ functions $\widetilde{p}(z, \bar\sigma) = \widetilde{p}(z)$, $\bar\sigma \geq 0$, defined as above. First of all, we remark that the roots of $z \mapsto \widetilde{p}(z, \bar\sigma)$ in $\mathbb{C}$ depend on $\bar\sigma$ continuously. The proof of this fact follows exactly like in the classical proof of the dependence of the continuity of the roots of a polynomial on its coefficients [Ort90, Theorem 3.1.1].

Lemma C.2. *Let $\bar\sigma_0 > 0$, and let z_i^0, $i = 0, \dots, N$, be the roots of $\widetilde p(z, \bar\sigma_0) = tz^2\sqrt{z^2+1} - 2Lz^3 + 2L\frac{\bar\sigma_0}{\lambda}$.*

(1) $z_k^0 \notin i\mathbb{R}$, $k = 0, \dots, N$.

(2) There exists $\theta_0 > 0$, s.t. for all $0 < \theta < \theta_0$, $\inf\limits_{z\in\theta+i\mathbb{R}} |p(z, \bar\sigma_0)| > 0$.

(3) There exists $\varepsilon_0 > 0$, s.t. for each $0 < \varepsilon < \varepsilon_0$ there exists $\delta > 0$, s.t. the roots z_i of $\widetilde p(z, \bar\sigma_0 + \delta)$ can be ordered to satisfy:

$$|z_i - z_i^0| < \varepsilon.$$

Proof. Proof of (1). For $\bar\sigma_0 > 0$, $z \mapsto \widetilde p(z, \bar\sigma_0)$ does not vanish on the imaginary axis. Indeed, for $z = i\omega$ with $|\omega| < 1$, we have that $\mathrm{Im}\widetilde p(z, \bar\sigma_0) = 2L\omega^3$, which vanishes only in 0, however, $\widetilde p(0, \bar\sigma_0) = 2L\frac{\bar\sigma_0}{\lambda} \neq 0$. For $|\omega| \geq 1$, we have that $\mathrm{Re}\widetilde p(z, \bar\sigma_0) = 2L\frac{\bar\sigma_0}{\lambda} \neq 0$.

Proof of (2). First of all, for $|z| \to +\infty$, $\widetilde p(z, \bar\sigma_0) = (t - 2L)z^3 + o(z^3)$, where $t > 2L$. From this and the fact that $\widetilde p$ does not vanish on the imaginary axis, it follows that there exists $\theta_0 > 0$, s.t., for all $0 < \theta < \theta_0$,

$$\inf_{z\in i\mathbb{R}\pm\theta} |p(z, \bar\sigma_0)| = \min_{z\in i\mathbb{R}\pm\theta} |p(z, \bar\sigma_0)| > \nu_0 > 0. \tag{47}$$

Proof of (3). Remark that $\widetilde p(z, \bar\sigma)$ has at most six roots: indeed, any root of $\widetilde p(z, \bar\sigma)$ solves the polynomial equation of the 6th degree:

$$\left(tz^2\sqrt{z^2+1}\right)^2 = \left(2Lz^3 - 2L\frac{\bar\sigma}{\lambda}\right)^2. \tag{48}$$

Next, let us fix $\varepsilon > 0$ sufficiently small so that $B(z_i^0, \varepsilon) \cap B(z_j^0, \varepsilon) = \emptyset$ for $i \neq j$, and $B(z_i^0, \varepsilon) \cap i\mathbb{R} = \emptyset$, for all i. Let

$$\nu := \min_i \min_{z\in\partial B(z_i^0,\varepsilon)} |\widetilde p(z, \bar\sigma_0)|.$$

Evidently, $\nu > 0$ by continuity of $\widetilde p$.

- remark that $|\widetilde p(z, \bar\sigma_0) - p(z, \bar\sigma)| = 2L\frac{|\bar\sigma_0 - \bar\sigma|}{\lambda}$. If $2L\frac{|\bar\sigma_0 - \bar\sigma|}{\lambda} < \nu$, then $|\widetilde p(z, \bar\sigma_0) - p(z, \bar\sigma)| < |\widetilde p(z, \bar\sigma_0)|$ on $\partial B(z_i^0, \varepsilon)$, and, by Rouché's theorem applied to $B(z_i^0, \varepsilon)$, the analytic in $\mathbb{C} \setminus i\mathbb{R}$ function $\widetilde p(z, \bar\sigma)$ has one root inside $B(z_i^0, \varepsilon)$.
- to show that all the roots of $\widetilde p(z, \bar\sigma)$ are given by perturbations of roots $\widetilde p(z, \bar\sigma_0)$, we remark that Rouché's theorem can be applied in the following setting. For any $R > 0$ sufficiently large we have

  $$\inf_{|z|=R} |\widetilde p(z, \bar\sigma_0)| > \nu_1 > 0.$$

  Then the above reasoning can be applied by taking $2L\frac{|\bar\sigma_0 - \bar\sigma|}{\lambda} < \min(\nu, \nu_0, \nu_1)$ (ν_0 as in (47)) and remarking that on

  $$\Gamma_R^+ = \{z \in i\mathbb{R} + \theta, |z| < R\} \cup \{|z| = R, \mathrm{Re}z > \theta\} \cup \cup_{i:\,\mathrm{Re}z_i^0>0} \partial B(z_i^0, \varepsilon),$$

  we again have $|\widetilde p(z, \bar\sigma) - \widetilde p(z, \bar\sigma_0)| < |\widetilde p(z, \bar\sigma_0)|$, and therefore both $\widetilde p(z, \bar\sigma_0)$ and $\widetilde p(z, \bar\sigma)$ do not have roots in the domain $\{z : \mathrm{Re}z > \theta, |z| < R\} \setminus \cup_{i:\,\mathrm{Re}z_i^0>0} B(z_i^0, \varepsilon)$. Because the above is valid for any sufficiently large R and sufficiently small θ, we conclude that the result holds true for $\mathbb{C}^+ \setminus \cup_{i:\,\mathrm{Re}z_i^0>0} B(z_i^0, \varepsilon)$.

  Similar reasoning applies to $\mathbb{C}^- \setminus \cup_{i:\,\mathrm{Re}z_i^0<0} B(z_i^0, \varepsilon)$.

□

Besides the fact that the roots of $\widetilde p$ for $\bar\sigma > 0$ are located in $\mathbb{C} \setminus i\mathbb{R}$, the above result also means that the number of roots of $\widetilde p(z)$ in the right-half is the same for any $\bar\sigma > 0$ (by continuity of the roots of the function $z \mapsto \widetilde p(z, \bar\sigma)$ on $\bar\sigma$ in $\mathbb{C} \setminus i\mathbb{R}$, changing $\bar\sigma$ would lead to roots getting close to the imaginary axis, which is

impossible as shown above). Therefore, it suffices to study the distribution of the roots of $z \mapsto \widetilde{p}(z, \bar{\sigma})$ for $\bar{\sigma} \to 0$. Remark that a root z of (48) is a root of $\widetilde{p}(z, \bar{\sigma})$ if and only if the following additional condition holds true (which stems from the chosen branch of $\sqrt{z}$)

$$\mathrm{Re}\left(\frac{1}{tz^2}\left(2Lz^3 - 2L\frac{\bar{\sigma}}{\lambda}\right)\right) > 0. \tag{49}$$

The polynomial equation (48) can be reduced to

$$P(z, \bar{\sigma}) := t^2 z^4 (z^2 + 1) - \left(2Lz^3 - 2L\frac{\bar{\sigma}}{\lambda}\right)^2 = 0. \tag{50}$$

For $\bar{\sigma} = 0$ it has a root $z_0^0 = 0$ of multiplicity 4 and two simple roots $z_\pm^0 = \pm i(1 - \frac{4L^2}{t^2})^{-1/2}$.

The above can be formalized as follows. Denoting by $\mathcal{Z}(p, \bar{\sigma})$ the set of the roots of $p(., \bar{\sigma})$ and by $\mathcal{Z}(P, \bar{\sigma})$ the set of the roots of $P(., \bar{\sigma})$, we have that

$$\mathcal{Z}(p, \bar{\sigma}) = \{z \in \mathcal{Z}(P, \bar{\sigma}) : (49) \text{ holds true}\}, \quad \bar{\sigma} > 0.$$

Because

$$\mathcal{Z}(P, 0) = \{0\} \cup \{z_\pm^0\},$$

for small $\bar{\sigma}$, the roots of $\mathcal{Z}(p, \bar{\sigma})$ are given by perturbations of $z_\pm^0$ and 0, as summarized in the lemma below.

Lemma C.3. *For all $\varepsilon > 0$, there exists $\delta > 0$, s.t. for all $0 < \bar{\sigma} < \delta$,*

$$\mathcal{Z}(p, \bar{\sigma}) \subset B(0, \varepsilon) \cup B(z_+^0, \varepsilon) \cup B(z_-^0, \varepsilon). \tag{51}$$

Let us now study the behaviour of the roots close to $z_\pm^0$.

Lemma C.4 ('Large' roots of $\widetilde{p}$)**.** *For $\bar{\sigma} \to 0$, two of the roots of $\widetilde{p}$ are given by*

$$z_\pm^{\bar{\sigma}} = z_\pm^0 + \frac{4L^2}{\lambda t^2}\bar{\sigma} + o(\bar{\sigma}), \quad \bar{\sigma} \to 0.$$

Proof. Let us study the roots $z_\pm^{\bar{\sigma}}$ of $\mathcal{Z}(P, \bar{\sigma})$ close to $z_\pm^0$. We have that

$$\partial_z P(z_\pm^0, 0) = \pm 2it^2(1 - \frac{4L^2}{t^2})^{-3/2}, \quad \partial_{\bar{\sigma}} P(z_\pm^0, 0) = \mp i\frac{8L^2}{\lambda}(1 - \frac{4L^2}{t^2})^{-3/2}.$$

By the implicit function theorem we then have that

$$z_\pm^{\bar{\sigma}} = z_\pm^0 + \frac{4L^2}{\lambda t^2}\bar{\sigma} + o(\bar{\sigma}), \quad \bar{\sigma} \to 0.$$

It remains to verify that these roots are indeed roots of $\widetilde{p}(z, \bar{\sigma})$, i.e. the condition (49). It automatically holds true, since it can be rewritten as follows:

$$\mathrm{Re}\left(\frac{1}{t(z_\pm^{\bar{\sigma}})^2}\left(2L(z_\pm^{\bar{\sigma}})^3 - 2L\frac{\bar{\sigma}}{\lambda}\right)\right) = \frac{2L}{t}\mathrm{Re}(z_\pm^{\bar{\sigma}}) - 2L\frac{\bar{\sigma}}{\lambda t(z_\pm^{\bar{\sigma}})^2} = \bar{\sigma}\frac{2L}{t}\frac{4L^2}{\lambda t^2} + 2L\frac{\bar{\sigma}}{\lambda|z_\pm^0|^2} + o(\bar{\sigma}),$$

which is indeed positive for $\bar{\sigma}$ sufficiently small. □

The above result shows in particular that for sufficiently small $\bar{\sigma}$ at least two roots of $\widetilde{p}$ are located in $\mathbb{C}^+$.

Lemma C.5 ('Small' roots of $\widetilde{p}$)**.** *For $\bar{\sigma} \to 0$, the function $\widetilde{p}$ has two roots in the vicinity of zero. Moreover, they have a strictly negative real part for $\bar{\sigma} > 0$.*

Proof. The roots of $\widetilde{p}(z,\bar{\sigma})$ satisfy

$$\widetilde{p}(z,\bar{\sigma}) = tz^2\sqrt{z^2+1} - 2Lz^3 + \frac{2L}{\lambda}\bar{\sigma} = 0.$$

We denote the last term of the above by ε and introduce the corresponding polynomial $\widetilde{p}_\varepsilon(z)$. Remark that for all $\varepsilon > 0$ sufficiently small there exists $\delta > 0$, $\delta < C\sqrt{\varepsilon}$ with some $C > 0$, s.t. $|tz^2\sqrt{z^2+1} - 2Lz^3 + \varepsilon - tz^2| < t|z|^2$ for all $|z| = \delta$. Therefore by Rouché's theorem, $\widetilde{p}_\varepsilon(z) = tz^2\sqrt{z^2+1} - 2Lz^3 + \varepsilon$ and $\widetilde{p}_\varepsilon^{(1)}(z) = tz^2$ have the same number of roots inside $B(0,\delta)$ (i.e. two roots).

It remains to prove that these two roots have a negative real part.

Let us now introduce $\widetilde{z} = \varepsilon^{-1/2}z$. Then the roots of $\widetilde{p}_\varepsilon(z)$ are (scaled) roots of

$$\widetilde{p}_\varepsilon^{(2)}(\widetilde{z}) = t\widetilde{z}^2\sqrt{1+\varepsilon\widetilde{z}^2} + 1 - 2L\varepsilon^{1/2}\widetilde{z}^3 = 0.$$

After setting $\eta = \varepsilon^{1/2}$, we can define $\widetilde{p}_\eta^{(3)} = \widetilde{p}_{\eta^2}^{(2)}$, i.e.

$$\widetilde{p}_\eta^{(3)}(\widetilde{z}) = t\widetilde{z}^2\sqrt{1+\eta^2\widetilde{z}^2} + 1 - 2L\eta\widetilde{z}^3 = 0.$$

The polynomial $\widetilde{p}_0^{(3)}(\widetilde{z})$ has two simple roots $\widetilde{z}_\pm^0 = \pm i/t$. It remains to use an implicit function theorem to obtain an expansion of the roots of $\widetilde{p}_\eta^{(3)}(\widetilde{z})$ as $\eta \to 0$:

$$\widetilde{z}_\pm^\eta = \pm i/t - \eta L/t^3 + o(\eta),$$

and this translates into the following expansion of the roots of $\widetilde{p}_\varepsilon(z)$ w.r.t. a small parameter ε:

$$z_\pm^\varepsilon = \varepsilon^{1/2}\left(\pm i/t - L/t^3\varepsilon^{1/2} + o(\varepsilon^{1/2})\right).$$

□

We finally can prove Proposition C.1.

Proof. As argued before, the number of roots of $p(z,\bar{\sigma})$ in $\mathbb{C}_*^+$ (resp. $\mathbb{C}_*^-$) remains constant for all $\bar{\sigma} > 0$. Lemmas C.3, C.4 and C.5 show that for small $\bar{\sigma} > 0$ $p(z,\bar{\sigma})$ has two roots in $\mathbb{C}_*^+$ and two roots in $\mathbb{C}_*^-$. The roots are complex conjugate since $p(\bar{z}) = \overline{p(z)}$. □