Optimizing Perfectly Matched Layers in Discrete Contexts

A. Modave¹*, E. Delhez², C. Geuzaine¹

¹ Department of Electrical Engineering and Computer Science (Montefiore Institute), University of Liège, Belgium
² Department of Aerospace and Mechanical Engineering (A&M), University of Liège, Belgium

SUMMARY

Perfectly Matched Layers (PMLs) are widely used for the numerical simulation of wave-like problems defined on large or infinite spatial domains. However, for both the time-dependent and the time-harmonic cases, their performance critically depends on the so-called absorption function. This paper deals with the choice of this function when classical numerical methods are used (based on finite differences, finite volumes, continuous finite elements and discontinuous finite elements). After reviewing the properties of the PMLs at the continuous level, we analyse how they are altered by the different spatial discretizations. In the light of these results, different shapes of absorption function are optimized and compared by means of both one- and two-dimensional representative time-dependent cases. This study highlights the advantages of the so-called shifted hyperbolic function, which is efficient in all cases and does not require the tuning of a free parameter, by contrast with the widely used polynomial functions. Copyright © 2014 John Wiley & Sons, Ltd.

Received …

KEY WORDS: perfectly matched layer; absorbing layer; absorption function; optimization; finite difference; finite volume; finite element

1. INTRODUCTION

The numerical simulation of wave propagation problems set in very large or infinite spatial domains remains a major challenge. A classical strategy consists in truncating the spatial domain to diminish the computational cost without altering too much the original solution. The artificial boundary of the truncated domain then requires a specific treatment allowing outgoing waves to leave the domain without spurious reflection. To this aim, numerous artificial boundary conditions, absorbing layers and special numerical techniques have been developed (see e.g. [2, 10, 14, 19, 21, 22, 23, 46] and references therein). Among them, the perfectly matched layer (PML) of Bérenger [5] exhibits features that have motivated its adaptation and its use in a wide range of cases from acoustics [28, 29], electromagnetism [1, 6, 31, 40], geophysical fluid dynamics [3, 33, 37], elastodynamics [4, 44, 45] and quantum mechanics [38]. Basically, the truncated domain is extended beyond the artificial boundary with an artificial layer where the original governing equations are transformed. When the PML techniques are used, outgoing waves are perfectly transmitted from the truncated domain to the PML whatever the angle of incidence – this is the perfect matching – and are then damped. The decay of the outgoing waves is controlled by an additional parameter: the absorption function \( \sigma(x) \). Under a specific condition on \( \sigma(x) \), the outgoing waves are completely damped and every spurious reflection is then avoided (see e.g. [5, 7]), which is the goal to achieve.

Unfortunately, the properties of the PMLs are altered when the problem is discretized [11, 47]. At the discrete level, the perfect matching is lost and, in the worst cases, the PML becomes
highly reflective. Spurious reflection of outgoing waves can hopefully be reduced by adjusting the discretization and the PML parameters. To this aim, very thick layers, mesh adaptation [34], modified discrete schemes [17, 20] and optimized absorption functions [11, 47] can all be used. However, the first two approaches imply an increase of the computational cost, and the others are case-dependent. In particular, polynomial functions are commonly used as absorption functions in order to ensure a progressive damping of the outgoing waves in the layer (see e.g. [6, 18]). These functions involve free parameters that are generally optimized by means of costly computational procedures or a priori error estimates [13, 16, 25, 38, 41]. Unfortunately, the optimum parameters depend on both the equations and the numerical scheme considered.

An alternative approach is offered by two promising absorption functions: the hyperbolic function \( \sigma_{\text{h}}(x) \) and the shifted hyperbolic function \( \sigma_{\text{sh}}(x) \). With them, a PML provides an exact treatment of the artificial boundary at the continuous level. At the discrete level, these functions do not require any tuning since the optimum value of their free parameter possesses a physical interpretation. This result is of high interest in industrial cases where the optimization of a free parameter is computationally very expensive, and sometimes downright impossible. This has been highlighted in several works [7, 8, 35, 36, 42] for a few wave-like problems and numerical schemes. However, conflicting conclusions about the relative performances of \( \sigma_{\text{h}} \) and \( \sigma_{\text{sh}} \) can be found in the literature.

In the specific context of the Helmholtz equation with a continuous finite element scheme, Bermúdez et al. [8] showed that \( \sigma_{\text{h}} \) outperforms both \( \sigma_{\text{sh}} \) and classical optimized polynomial functions. By contrast, for a time-dependent first-order wave-like system with a finite difference scheme, we showed [35] that \( \sigma_{\text{sh}} \) performs very well, while \( \sigma_{\text{h}} \) provides bad results.

This paper deals with the optimization of PMLs in discrete contexts for wave-like problems based on first-order systems. In order to reach conclusions as case-independent as possible, all simulations are systematically performed with different numerical methods, and interpretations are proposed. Since the absorption function is a key parameter of this optimization problem, a large part of the paper is dedicated to its study. In particular, we confirm that the functions \( \sigma_{\text{h}} \) and \( \sigma_{\text{sh}} \) do not require any tuning whatever the numerical method. In all cases, \( \sigma_{\text{sh}} \) provides good results that are at least comparable to those obtained with the optimized polynomial functions, while \( \sigma_{\text{h}} \) is less robust.

This paper is organized as follows. The PML equations are derived for a fundamental case in Section 2. Classical results on PMLs at the continuous level are reviewed in Section 3. In Section 4, we propose a preliminary analysis of the PML in one-dimensional discrete contexts. The change of properties of the discretized PMLs and the influence of the problem parameters on their effectiveness are then highlighted. In particular, the specific role of the absorption function \( \sigma(x) \) is shown and interpreted, together with the existence of optimum values in discrete contexts. In Section 5, the polynomial and hyperbolic absorption functions are optimized and compared by means of benchmarks involving waves with normal and oblique incidences. Thereby, an interpretation of the results obtained with the hyperbolic functions is proposed.

2. THE PERFECTLY MATCHED LAYER

Let us consider a general problem with non-dispersive scalar waves, defined on the unbounded domain \( \mathbb{R}^d \), where \( d \) is the dimension of the space. The time-evolution of a scalar field \( p(x, t) \) and a vector field \( u(x, t) \) is governed by the first-order system

\[
\begin{align*}
\frac{\partial p}{\partial t} + a \nabla \cdot u &= 0, \\
\frac{\partial u}{\partial t} + b \nabla p &= 0,
\end{align*}
\]

where \( a \) and \( b \) are physical parameters, \( x \) denotes the spatial position and \( t \) is the time. Initial conditions on both fields close the differential problem. This problem corresponds to a basic case encountered in various physical contexts.

The PML equations are introduced for a modified version of this problem that is defined on the truncated domain \( \Omega = \{ x \in \mathbb{R}^d : x < 0 \} \) extended with the layer \( \Sigma = \{ x \in \mathbb{R}^d : x \in [0, \delta] \} \), where
x is the Cartesian coordinate corresponding to the direction $e_x$ and $\delta$ denotes the thickness of the layer. Most PMLs can be interpreted as media where the coordinates are stretched in the frequency space, as firstly introduced by Chew and Weedon [12]. Following this idea, the PML equations are built by considering the time-harmonic form of the system (1), i.e.

$$
\begin{align*}
-\omega \hat{p} + a \nabla \cdot \hat{u} &= 0, \\
-\omega \hat{u} + b \nabla \hat{p} &= 0,
\end{align*}
$$

(2)

where $\hat{p}$ and $\hat{u}$ are the time-harmonic fields and $\omega$ is the angular frequency. The real coordinate $x$ is then stretched using the complex mapping

$$
x \to x^* = x - \frac{1}{\omega} \int_0^x \sigma(x') \mathrm{d}x',
$$

(3)

where $\sigma(x)$ is the absorption function, which is positive in $\Sigma$ and equal to zero inside the domain $\Omega$. Basically, the original variable $x$ is replaced with $x^*$ in the differential operators of the system (2). The resulting system is then rewritten in terms of the physical variable $x$ thanks to the change of variable $x^*(x)$. The time-harmonic system then becomes

$$
\begin{align*}
-\omega \hat{p} + a \nabla \cdot \hat{u} &= -\sigma \hat{u}_x, \\
-\omega \hat{u} + b \nabla \hat{p} &= -\sigma \hat{p}_x e_x,
\end{align*}
$$

where $\hat{u}_x$ is the $x$-component of $\hat{u}$ and $e_x$ indicates the $x$-direction. The time-dependent system is finally obtained by defining an additional equation and applying the inverse Fourier-transform:

$$
\begin{align*}
\frac{\partial p}{\partial t} + a \frac{\partial u}{\partial x} &= -\sigma q, \\
\frac{\partial u}{\partial t} + b \frac{\partial p}{\partial x} &= -\sigma u, \\
\frac{\partial q}{\partial t} + a \frac{\partial u}{\partial x} &= -\sigma q,
\end{align*}
$$

(4)

where $q(x, t)$ is an additional field, which is defined only in $\Sigma$. At the interface between $\Omega$ and $\Sigma$, the continuity of the fields is prescribed. We choose to use a homogeneous Dirichlet condition on $u$ for the external boundary, i.e. $u \cdot e_x = 0$ at $x = \delta$. Note that the exact nature of this boundary condition has little influence on the solution if the outgoing waves are efficiently damped in the PML [39].

In the fundamental case where all waves propagate along the normal to the artificial boundary, the problem becomes one-dimensional. In both the one-dimensional domain $\Omega = \mathbb{R}^-$ and the layer $\Sigma = [0, \delta]$, the scalar fields $p(x, t)$ and $u(x, t)$ are then governed by

$$
\begin{align*}
\frac{\partial p}{\partial t} + a \frac{\partial u}{\partial x} &= -\sigma p, \\
\frac{\partial u}{\partial t} + b \frac{\partial p}{\partial x} &= -\sigma u.
\end{align*}
$$

This system is an important asymptotic case for all first-order PML systems of the literature. The scope of the discussion based on it is therefore very large, and does not only apply for problems with non-dispersive scalar waves.

3. KEY PROPERTIES IN THE CONTINUOUS CONTEXT

At the continuous level, the PML has interesting properties that make it an attractive method to deal with the truncation of a large or infinite domain. These are reviewed and demonstrated in this section by means of a classical plane-wave analysis (see e.g. [5, 6, 9]).
The original system (1) supports time-harmonic plane-wave solutions of the form

$$\begin{align*}
\begin{pmatrix} p(x, t) \\ u(x, t) \end{pmatrix} &= \begin{pmatrix} P \\ U \end{pmatrix} e^{i(k \cdot x - \omega t)},
\end{align*}$$

(5)

where $P$ and $U$ are the amplitudes and $k$ is the wave vector, which are related by the dispersion relation

$$\omega = \sqrt{ab} \|k\|$$

and the amplitude relation

$$P \frac{k}{\|k\|} = \sqrt{\frac{a}{b}} U.$$  

(6)

These waves correspond to a non-dispersive and non-dissipative medium. In the PML, the properties of the waves are modified due to the complex coordinate stretch (3). The travelling modes (5) indeed become damped waves shaped as

$$\begin{align*}
\begin{pmatrix} p(x, t) \\ u(x, t) \end{pmatrix} &= \begin{pmatrix} P \\ U \end{pmatrix} e^{i(k \cdot x - \omega t)} e^{-\gamma(x)},
\end{align*}$$

(7)

with the attenuation factor

$$\gamma(x) = \frac{\cos \theta}{\sqrt{ab}} \int_0^x \sigma(x') \, dx', \quad \forall x \in [0, \delta],$$

where $\theta = \arccos(k_x/\|k\|)$ is the angle of incidence. Both the dispersion relation and the amplitude relation are identical to those of the original medium. Therefore, plane waves propagating in the PML medium have exactly the same properties as those of the original medium, but with an additional damping factor (the second exponential of the solution (7)). The PML medium is both non-dispersive and dissipative. Note that the shape of the wave decay is independent of the frequency, but not of the angle of incidence.

The interface between the domain $\Omega$ and the layer $\Sigma$ is transparent to all travelling waves. This can be shown by considering a representative problem where the domain $\Omega$ is extended with a semi-infinite layer, i.e. with $\delta \to +\infty$. Let us assume that an incident wave propagating in $\Omega$ towards the layer splits into a reflected part (in $\Omega$) and a transmitted part (in $\Sigma$). The solution then reads

$$\begin{align*}
p(x, t) &= \begin{cases} 
p^r e^{i(k \cdot x - \omega t)} + P r e^{-2ik_x x} e^{i(k \cdot x - \omega t)}, & \forall x \in \Omega, \\
p^\ell e^{i(k \cdot x - \omega t)} e^{-\gamma(x)}, & \forall x \in \Sigma, \end{cases}
\end{align*}$$

(8)

$$\begin{align*}
u(x, t) &= \begin{cases} 
u^r e^{i(k \cdot x - \omega t)} + U r e^{-2ik_x x} e^{i(k \cdot x - \omega t)}, & \forall x \in \Omega, \\
u^\ell e^{i(k \cdot x - \omega t)} e^{-\gamma(x)}, & \forall x \in \Sigma. \end{cases}
\end{align*}$$

(9)

Injecting this solution in the interface conditions and using the amplitude relation (6), one immediately obtains that the amplitudes of the reflected wave, $P^r$ and $U^r$, are zero whatever the value of $\sigma$. Therefore, the reflection coefficient associated with the interface is zero, i.e.

$$r_{\text{interf}} = \left| \frac{P^r}{P^r} \right| = 0.$$  

(10)

Every incident wave is perfectly transmitted from the truncated domain to the layer whatever its frequency and angle of incidence. Therefore the interface is perfectly matched.

If the layer thickness $\delta$ is finite, the outgoing waves that are not sufficiently damped in $\Sigma$ can be reflected at the external boundary (i.e. at $x = \delta$) and come back in $\Omega$. This is shown by assuming the plane-wave solution has an additional component that is reflected by the external boundary, i.e.

$$\begin{align*}
p(x, t) &= \begin{cases} 
p^r e^{i(k \cdot x - \omega t)} + P r e^{-2ik_x x} e^{i(k \cdot x - \omega t)}, & \forall x \in \Omega, \\
p^l e^{i(k \cdot x - \omega t)} e^{-\gamma(x)} + P b e^{-2ik_x x} e^{i(k \cdot x - \omega t)} e^{\gamma(x)}, & \forall x \in \Sigma, \end{cases}
\end{align*}$$

(11)

$$\begin{align*}
u(x, t) &= \begin{cases} 
u^r e^{i(k \cdot x - \omega t)} + U r e^{-2ik_x x} e^{i(k \cdot x - \omega t)}, & \forall x \in \Omega, \\
u^l e^{i(k \cdot x - \omega t)} e^{-\gamma(x)} + U b e^{-2ik_x x} e^{i(k \cdot x - \omega t)} e^{\gamma(x)}, & \forall x \in \Sigma. \end{cases}
\end{align*}$$

(12)
Injecting this solution in the boundary condition $e_x \cdot u = 0$ at $x = \delta$ and using the perfect matching and the amplitude relation, one obtains the reflection coefficient associated with a finite layer, i.e.

$$r_{\text{pml}} = \left| \frac{P_r}{P_i} \right| = \exp \left[ -2 \cos \frac{\theta}{\sqrt{ab}} \int_0^\delta \sigma(x') \, dx' \right]. \quad (13)$$

This coefficient gives an indication on the reflection of outgoing waves by the layer. In the particular case where the absorption function $\sigma(x)$ has an unbounded integral, i.e.

$$\int_0^\delta \sigma(x') \, dx' = +\infty, \quad (14)$$

the reflection coefficient is zero. There is then no reflection and the layer is perfectly absorbing.

4. PRELIMINARY ANALYSIS IN DISCRETE CONTEXTS

When using a PML together with a numerical scheme, it is worthwhile to preserve its ability to accurately simulate the artificial boundary of the truncated domain. This can be studied from two complementary points of view.

1. The different contributions to the global error on the solution can be identified and quantified separately: the modeling error, caused by the replacement of the original continuous problem (defined on $\mathbb{R}^d$) with the modified version (defined on $\Omega$), and the numerical error, generated by the discretization of the modified continuous problem. The former can be quantified with the reflection coefficient $r_{\text{pml}}$ (13) or a $L^2$-error on the solution. For a finite difference scheme, the latter can be evaluated using the truncation error of the scheme.

2. The discrete problem can also be considered itself as a wave-like problem, where the discrete equations support discrete plane-wave solutions. The purpose of a discretized PML is then to absorb as accurately as possible the discrete outgoing waves. The challenge consists here in defining and computing discrete versions of the reflection coefficient.

These points of view lead to different approaches to optimize PMLs: minimizing both modeling and numerical errors [38], and minimizing a discrete reflection coefficient [13, 25, 41]. The former way has the advantage of highlighting the different sources of error, while the latter provides a single quantity to minimize. In this section, the properties of the PML are analysed in discrete contexts taking the second point of view. Discrete reflection coefficients are first provided by a discrete time-harmonic plane-wave analysis in a basic one-dimensional finite difference case. Then, our approach is validated and extended to other numerical methods by using simulations in the time domain.

4.1. Discrete plane-wave analysis – Finite difference scheme

Let us consider a spatial discretization of the problem in a one-dimensional case with a finite difference (FD) method. The semi-discrete fields $\tilde{\rho}_{i+1/2}(t)$ and $\tilde{u}_i(t)$ take their values on staggered regular grids, at the discrete points $x_{i+1/2} = (i + 1/2)\Delta x$ and $x_i = i\Delta x$, respectively, where $i$ is the spatial index and $\Delta x$ is the spatial step. Using central finite differences, the governing semi-discrete equations read

$$\begin{cases} 
\frac{d\tilde{\rho}_{i+1/2}}{dt} + a \frac{\tilde{u}_{i+1} - \tilde{u}_i}{\Delta x} = -\sigma_{i+1/2} \tilde{\rho}_{i+1/2}, \\
\frac{d\tilde{u}_i}{dt} + b \frac{\tilde{\rho}_{i+1/2} - \tilde{\rho}_{i-1/2}}{\Delta x} = -\sigma_i \tilde{u}_i.
\end{cases} \quad (15)$$

The discrete values of the absorption function, $\sigma_i$ and $\sigma_{i+1/2}$, are equal to zero in the truncated domain ($i < 0$) and are positive in the PML ($i \geq 0$).
A. MODAVE ET AL.

Spatial coordinate $x$

Plane-wave solution

Domain $\Omega$

Layer $\Sigma$

$\varepsilon = c/\sigma$

Figure 1. Illustration of the exponential decay of a plane wave in a one-dimensional PML for a constant absorption function $\sigma(x) = \bar{\sigma}$. The characteristic length of the decay is $c/\bar{\sigma}$.

The semi-discrete system (15) supports plane-wave solutions similar to the continuous systems, but with slightly different properties. In the domain $\Omega$, where the source terms are removed, the elementary solution reads

$$\tilde{p}_{i+1/2}(t) = P e^{i(k(i+1/2)\Delta x - \omega t)},$$

(16)

$$\tilde{u}_i(t) = U e^{i(ki\Delta x - \omega t)},$$

(17)

and the wave number $k$ (real) and the angular frequency $\omega$ (real and positive) are now related to each other by the dispersion relation

$$k = \pm \frac{2}{\Delta x} \arcsin \left( \frac{\Delta x}{2} \frac{\omega}{c} \right),$$

(18)

with $c = \sqrt{ab}$. Since the phase velocity $\omega/|k|$ depends on the wave number, these plane waves are dispersive, contrary to their continuous versions. The non-dispersive behavior of continuous waves is recovered when the spatial step $\Delta x$ is small with respect to the wavelength $c/\omega$. This means that the discrete lattice must be fine enough to be able to capture the oscillatory behavior of waves. In the PML $\Sigma$, the oscillatory behavior of discrete plane waves is altered by the source terms, and the decay of these waves varies with the frequency, in contrast with the continuous case. If the absorption function is constant ($\sigma_i = \sigma_{i+1/2} = \bar{\sigma}, \forall i \geq 0$), the elementary solution of the system can be written as a damped plane wave

$$\tilde{p}_{i+1/2}(t) = P e^{i(\beta(i+1/2)\Delta x - \omega t)},$$

(19)

$$\tilde{u}_i(t) = U e^{i(\beta_i\Delta x - \omega t)},$$

(20)

where $\beta$ is a complex number that is linked to the problem parameters by the dispersion relation

$$\beta = \pm \frac{2}{\Delta x} \arcsin \left( \frac{\Delta x}{2} \frac{\omega + i\bar{\sigma}}{c} \right).$$

(21)

The real part of $\beta$ corresponds to the wave number and its imaginary part gives the attenuation factor. To recover the continuous dispersion relation, the spatial step $\Delta x$ must be small with respect to both $c/\omega$ and $c/\bar{\sigma}$. The first condition is the same as in the domain $\Omega$ (with $\sigma = 0$). The second condition can be interpreted considering the shape of the solution in the layer, which is illustrated in Figure 1. For a constant absorption function, the decay of waves is exponential, with the characteristic length $c/\bar{\sigma}$. This decay is captured by the discrete mesh only if the spatial step is smaller than the characteristic length, i.e., $\Delta x \ll c/\bar{\sigma}$.
OPTIMIZING PERFECTLY MATCHED LAYERS IN DISCRETE CONTEXTS

The discrete versions of the reflection coefficients can be derived using the same procedure as in the continuous case (Section 3). Let us assume a semi-infinite truncated domain terminated with a PML of infinite or finite thickness, and semi-discrete solutions similar to the continuous ones (8)–(9) and (11)–(12). The discrete reflection coefficients are the ratio of the amplitudes of the incident and the reflected discrete plane waves. Injecting the semi-discrete solutions in the discrete scheme and using the discrete versions of both dispersion and amplitude relations, we get, respectively, the coefficient associated to the interface (or an infinite PML)

\[ r_{\text{interf}}^* = \left| \frac{e^{-i\beta \Delta x / 2} - e^{-i k \Delta x / 2}}{e^{-i\beta \Delta x / 2} + e^{ik \Delta x / 2}} \right| \]

and the one corresponding to the PML of thickness \( \delta \)

\[ r_{\text{pml}}^* = \left| \frac{\cos(\beta \delta + \beta \Delta x / 2) e^{-ik \Delta x / 2} - \sin(\beta \delta) e^{-ik \Delta x / 2}}{\cos(\beta \delta + \beta \Delta x / 2) e^{-ik \Delta x / 2} + \sin(\beta \delta) e^{ik \Delta x / 2}} \right|, \]

where \( k \) and \( \beta \) are given by the dispersion relations (18) and (21) with the plus sign in either case. The complete calculations are detailed in Appendix A.2.

Both continuous and discrete versions of the reflection coefficients are plotted as functions of \( \bar{\sigma} \) for a set of parameters in Figure 2. It is interesting to compare and interpret these curves for small and large values of \( \bar{\sigma} \) separately. For a small value of \( \bar{\sigma} \), the behaviour of the continuous solution is accurately reproduced by the numerical scheme. Indeed, when \( \bar{\sigma} < c/\delta \), the curves of the discrete reflection coefficients are close to those of the continuous ones. The outgoing waves are perfectly transmitted from the domain to the PML. However, they are not sufficiently damped, and are then reflected by the external boundary and come back in the domain. This undesirable behavior is due to the modeling error that is already present in the continuous model, and is reproduced by the discrete model. If \( \bar{\sigma} \) is large, the interface domain/layer is reflective in the discrete context, while it is perfectly matched in the continuous one. This spurious reflection of waves is caused by the discretization error. When the characteristic length of the exponential decay of outgoing waves is smaller than the spatial step, i.e. \( c/\bar{\sigma} < \Delta x \), the behavior of the solution cannot be reproduced with the discrete lattice and large numerical errors appear. A constant absorption function \( \bar{\sigma} \) must therefore be chosen in such a way to damp enough outgoing waves without inducing a too sharp decrease of the fields. There exists an optimum value \( \bar{\sigma}_{\text{opt}} \) that corresponds to this compromise. It

Figure 2. Reflection coefficients provided by plane-wave analyses in both continuous (\( r_{\text{interf}} \) and \( r_{\text{pml}} \)) and discrete (\( r_{\text{interf}}^* \) and \( r_{\text{pml}}^* \)) contexts. The parameters are \( \delta = 5 \Delta x, \lambda = 13 \Delta x \) and \( \Delta x = 10^3 \). All quantities are normalized to both the phase velocity \( \sqrt{ab} \) and the impedance \( \sqrt{a/b} \), and are then dimensionless.
minimizes the coefficient (22), and is such that the characteristic length of the exponential decay \( c/\bar{\sigma}_{\text{opt}} \) is in the range \([\Delta x, \delta]\), as shown in Figure 2.

The impact of the spatial step \( \Delta x \) and the thickness of the layer \( \delta \) on the errors can be also interpreted from Figure 2. An increase of \( \delta \) moves the curve \( r_{\text{pml}}(\bar{\sigma}) \) to the left, and a decrease of \( \Delta x \) moves the curve \( r_{\text{interf}}(\bar{\sigma}) \) to the right. Therefore, for a given value of \( \bar{\sigma} \), this leads to a decrease of the modeling or numerical error, respectively. Both ways can potentially decrease the global error. However, they require an increase of the computational cost, and it is not certain that the global error will actually decrease in all cases. Indeed, an increase of \( \delta \) (resp. decrease of \( \Delta x \)) does not improve the layer if the global error is largely dominated by the numerical error (resp. modeling error).

The curves associated to the discrete reflection coefficients do not notably change with the wavelength \( \lambda \). In particular, as shown in Figure 3a, the optimum values \( \bar{\sigma}_{\text{opt}} \) remain in the range \([c/\delta, c/\Delta x]\), whatever the wavelength. The corresponding minimum values of the reflection coefficient do not vary significantly for large enough wavelength (Figure 3b). Therefore, the effectiveness of the PML is equivalent for all plane waves with nearly the same optimum value \( \bar{\sigma}_{\text{opt}} \), except when the wavelength is close to the spatial step. The formulas of the discrete reflection coefficients for the low-frequency asymptotic case are given at the end of Appendix A.2. It is a remarkable fact that these formulas are independent of the frequency, which confirms our analysis.

As a conclusion, \( \bar{\sigma} \), \( \delta \) and \( \Delta x \) are the key parameters that influence the effectiveness of the PML in this discrete finite difference context. For a constant absorption function, the modeling error is related to the value \( \delta \), while the numerical error obviously depends on \( \Delta x \). The most appropriate approach to improve a PML is to choose \( \delta \) and \( \Delta x \) according to the acceptable computational cost, and after to find the corresponding optimum value of \( \bar{\sigma} \). If the obtained global error is too large, the procedure must be done again with a larger \( \delta \) and/or a smaller \( \Delta x \), and thus an increase of the computational cost.

### 4.2. Time-dependent simulations – Extention to other numerical schemes

In order to validate our approach and to extend the discussion to other numerical methods, the properties of the PML are now studied with time-dependent numerical simulations. We consider numerical schemes based on the finite difference (FD) method, as well as finite volume (FV), stabilized continuous finite element (CG) and discontinuous finite element (DG) methods.
Figure 4. Illustration of the solution of the one-dimensional benchmark at the initial and final instants. A Gaussian-shaped pulse that moves to the right is prescribed by the initial condition. After a reflection by the layer, the reflected part of the pulse moves to the left, and is finally centered in the right part of the domain.

Convenience, all quantities are normalized to both the velocity $\sqrt{ab}$ and the impedance $\sqrt{a/b}$, and are then dimensionless.

The time-harmonic one-dimensional benchmark is adapted for the time domain with a transient incident signal. The finite truncated domain $\Omega = [-L, 0]$ is extended with the PML $\Sigma = [0, \delta]$. An incident Gaussian-shaped pulse is used as initial condition with

$$p(x, 0) = \exp \left( -\frac{(x + L/4)^2}{R^2} \right),$$
$$u(x, 0) = \exp \left( -\frac{(x + L/4)^2}{R^2} \right),$$

where $R$ is a constant parameter. This pulse covers a broad range of frequencies, in contrast to the incident sine wave of the time-harmonic benchmark. As illustrated in Figure 4, it is initially centered at the middle of $[-L/2, 0]$. The parameter $R$ is chosen such that both fields are initially negligible outside this region. As time goes by, the pulse moves towards the layer and is partly reflected. At the end of the simulation ($t_f = L/2$), the reflected part of the pulse is mainly in $[-L/2, 0]$. Due to the numerical dispersion of the discrete scheme, some numerical modes can reach $[-L, -L/2]$ and the shape of the reflected signal can be non-Gaussian. The boundary condition $u = 0$ is used at both $x = -L$ and $x = \delta$.

The effectiveness of the PML is now quantified using the relative error $\xi_r$ on the numerical solution defined by

$$\xi_r = \sqrt{\frac{E_{\text{pml}}(t_f)}{E_{\text{wall}}(t_f)}}, \quad (23)$$

where $E_{\text{pml}}(t_f)$ is the global energy associated to the numerical solution in $[-L, 0]$ at the end of the simulation, and $E_{\text{wall}}(t_f)$ is the one obtained by replacing the PML with the perfectly reflecting boundary condition $u = 0$ at $x = 0$. In this dimensionless context, the global energy $E$ is defined by

$$E(t_f) = \int_{-L}^{0} \left( \frac{1}{2} p^2(x, t_f) + \frac{1}{2} u^2(x, t_f) \right) dx.$$

The value of $\xi_r$ can be interpreted as the part of the global energy that is reflected by the layer. The value 0 corresponds to a perfectly absorbing layer, while 1 is for a perfectly reflective layer.
The discrete reflection coefficient \( r_{\text{pml}}^* \) and the relative error \( \xi_r \) are complementary ways of quantifying the effectiveness of a discretized PML. The former is for a particular frequency (time-harmonic context), while the latter directly accounts for a range of frequencies through the initial incident pulse (time-dependent context). In the particular case where a harmonic plane wave is used instead of the Gaussian-shaped incident pulse, the relative error reduces to the discrete reflection coefficient, \( i.e. \xi_r = r_{\text{pml}}^* \).

The relative error \( \xi_r \) has been computed for different values of the parameter \( \bar{\sigma} \) in \([10^{-6}, 10^{0}]\) with different numerical schemes in a unique setting. The curves of \( \xi_r \) as a function of \( \bar{\sigma} \) are plotted in Figure 5 (linear scale) and Figure 6 (logarithmic scale). The finite difference (FD) scheme described in Section 4.1 is considered, as well as a finite volume (FV) scheme and finite element schemes based on continuous and discontinuous Galerkin methods (CG and DG). Both CG and DG schemes are described in Appendix B, while the FV scheme corresponds to the DG scheme with piecewise constant basis functions. We use centered and upwind fluxes for FV and DG schemes, and first- and second-order polynomial basis functions for CG and DG schemes. In all cases, the spatial step is \( \Delta x = 10^3 \), the size of the domain is \( L = 500 \Delta x \), the thickness of the layer is \( \delta = 5 \Delta x \), and the parameter of the initial pulse is \( R = 10 \Delta x \). The time-stepping is performed using the Leapfrog scheme with \( \Delta t = 0.5 \times 10^3 \) (FD) or the Crank-Nicolson scheme with \( \Delta t = 2.5 \times 10^3 \) (FV, CG and DG). To stabilize the CG scheme, PSPG terms are added in the formulation (see Appendix B) with the numerical diffusion parameter \( \kappa = 10^3 \).

The numerical results obtained with the FD scheme are consistent with the plane-wave analysis of the previous section. Indeed, in Figure 5a, the curve of the relative error \( \xi_r(\bar{\sigma}) \) is close to the one of the discrete reflection coefficient \( r_{\text{pml}}^*(\bar{\sigma}) \) obtained with the plane-wave analysis. The small difference can be explained by the change of incident signal: a harmonic one with a single frequency for \( r_{\text{pml}}^* \) and a Gaussian-shaped one covering a range of frequencies for \( \xi_r \), including high frequencies.

For \( \bar{\sigma} < c/\delta \), the continuous solution, and then the corresponding modeling error, is accurately reproduced by all schemes. All relative error curves are indeed close to the one of the continuous reflection coefficient \( r_{\text{pml}} \). By contrast, for \( \bar{\sigma} > c/\delta \), \( i.e. \) when the exponential decay of outgoing waves is too sharp to be reproduced by the discrete mesh, the behavior of the discretized PML varies with the numerical scheme. With the CG scheme and the FV/DG schemes (with centered fluxes), the curves \( \xi_r(\bar{\sigma}) \) look like the one obtained with the FD scheme (Figures 5b and 5c): there is an optimum value of \( \bar{\sigma} \) and, beyond this value, the relative error increases with \( \bar{\sigma} \). For very high values of \( \bar{\sigma} \), the interface domain/layer becomes perfectly reflective. The optimum value of \( \bar{\sigma} \) increases when the order of the polynomial basis functions increases (Figures 6b and 6c). This can be explained by considering that a higher order method can more accurately reproduce the rapid variations of the solution. Therefore, a higher \( \bar{\sigma} \), which reduces the characteristic length of the exponential decay, can be used. With the FV/DG schemes and upwind fluxes, unlike with other schemes, the PML remains highly absorbing for large values of \( \bar{\sigma} \) (Figure 5d). This particular behavior is due to the upwind fluxes, which are based on an exact Riemann solver. At the interface, this solver naturally splits the parts of the solution that propagate inward and outward. The outgoing Riemann invariant, which contains outgoing waves, is correctly computed, while the incoming Riemann invariant is computed using the overdamped values of the fields of the layer, and is then close to zero. This corresponds to prescribing that there are no incoming waves. Prescribing the incoming Riemann invariant to zero is a common absorbing boundary condition (ABC) strategy, which is exact in one-dimensional cases. As shown in Figure 6d, the PML reaches the accuracy of this ABC for large values of \( \bar{\sigma} \).

Unfortunately, as shown later, this ideal behavior is lost for multidimensional problems with oblique incident waves. Indeed, the ABC is then approximate and becomes reflective. Similarly, the PML becomes reflective for too large values of \( \bar{\sigma} \).

All these numerical results corroborate the conclusion of the plane-wave analysis performed in the FD context, and extend it to other numerical methods. In nearly all cases, there is an optimum value of the constant absorption function \( \bar{\sigma} \) to use in discrete contexts. The important exception observed with the FV/DG schemes and upwind fluxes is a particular case that does not occur in multidimensional problems.
Figure 5. Relative errors $\xi_r$ of the one-dimensional benchmark as a function of the constant absorption function $\bar{\sigma}$ for different numerical methods. The continuous reflection coefficient $r_{\text{pml}}$ is plotted in each graph, as well as the discrete version $r_{\text{pml}}^*$ in FD case for $\lambda = 13 \Delta x$. All values are dimensionless.
Figure 6. Same as Figure 5, but with a logarithmic scale. In Figure 6d, horizontal lines are plotted at the levels corresponding to the relative errors obtained with an ABC prescribed at $x = 0$. 
OPTIMIZING PERFECTLY MATCHED LAYERS IN DISCRETE CONTEXTS

5. SPATIALLY VARYING ABSORPTION FUNCTIONS

Smoothly increasing functions are widely used as absorption functions. They induce a gradual decay of outgoing waves that is more easily captured by numerical methods than a sharp exponential decay. At the discrete level, the effectiveness of the PML is then easily improved at no additional cost. Polynomial absorption functions are most frequently used (see e.g. [18] and the references of the present paper), i.e.

\[ \sigma_n(x) = \alpha \left( \frac{x}{\delta} \right)^n, \]

where \( n \) and \( \alpha \) are positive parameters. The power \( n \) is generally equal to 2 or 3, which correspond respectively to parabolic \( \sigma_2 \) and cubic \( \sigma_3 \) functions. Similarly to the value \( \bar{\sigma} \) of constant absorption functions, the parameter \( \alpha \) of polynomial functions must be large enough to damp the outgoing waves, and small enough to avoid a too sharp variation of the solution that cannot be reproduced accurately by the numerical scheme. Nevertheless, the optimum \( \alpha \) is currently not clearly identified, and no general rule exists to choose it.

As an alternative, hyperbolic functions were proposed by Bermúdez et al. [8]:

\[ \sigma_h(x) = \frac{\alpha}{\delta - x}, \]
\[ \sigma_{sh}(x) = \frac{\alpha}{\delta - x} - \frac{\alpha}{\delta}, \]

where \( \alpha \) is an additional positive parameter to choose. The second function is called shifted hyperbolic function, because it is pushed down to ensure a zero value at the interface domain/layer (i.e. at \( x = 0 \)). Since both functions have an unbounded integral, the PML is theoretically perfectly absorbing at the continuous level (the condition (14) is met). In discrete contexts, the parameter \( \alpha \) must be optimized for the same reason as with other functions. Fortunately, both functions are close to optimum with \( \alpha \approx c \) (or \( \alpha \approx 1 \) in the dimensionless context). This was first observed by Bermúdez et al. [7] in a specific time-harmonic acoustic context with a continuous finite element scheme. This result is extended hereafter in time-dependent contexts with different numerical methods and interpretations are proposed.

This section deals with the choice of the spatially varying function to use as absorption function \( \sigma(x) \). Polynomial and hyperbolic functions are optimized and compared, and the optimum value of the free-parameter \( \alpha \) is discussed. To this aim, the one-dimensional benchmark is used again, as well as a two-dimensional benchmark that deals with waves having oblique incidences.

5.1. Optimization for normal waves

Consider again the one-dimensional benchmark of Section 4.2. The relative error \( \xi_r \) (23) is now computed with different values of the parameter \( \alpha \) for the different absorption functions and different numerical settings. We consider FD, CG and DG schemes, with first-order polynomial basis functions for the finite element schemes (CG and DG), and both centered and upwind fluxes for the DG scheme. In each case, two PML thicknesses \( \delta \) are considered, as well as two sets of numerical parameters: those of Section 4.2 (set A) and smaller ones (set B). The curves of the relative error \( \xi_r \) as a function of \( \alpha \) are shown in Figures 7, 8, 9 and 10.

The curves \( \xi_r(\alpha) \) exhibit the same behavior in all cases, except when the DG scheme is used with upwind fluxes (Figure 10). In this case, the result obtained with the constant absorption function is recovered: the PML is highly absorbing with very high value of \( \alpha \), and the value of the relative error is close to the one provided by the ABC. This particularity can be explained again by the specific definition of the upwind fluxes (see Section 4.2).

With the other numerical schemes, the curves \( \xi_r(\alpha) \) can present several minima. A systematic optimization of absorption functions must then be carefully done. For polynomial functions, the best \( \alpha \), which corresponds to the global minimum of each curve, is always in the range \( 10^{-3} - 10^{-2} \), except for the DG scheme with centered fluxes, where the best \( \alpha \) can have very different values:
(a) FD scheme with $\delta = 5 \times 10^3$ and set A

(b) FD scheme with $\delta = 10^4$ and set A

(c) FD scheme with $\delta = 5 \times 10^3$ and set B

(d) FD scheme with $\delta = 10^4$ and set B

Figure 7. One-dimensional benchmark with the finite difference (FD) scheme. Relative error $\xi_r$ as a function of the parameter $\alpha$ for different absorption functions $\sigma(x)$ and two PML thicknesses $\delta$. In each case, two sets of numerical parameters are considered: $\Delta x = 10^3$, $\Delta t = 500$ (set A) and $\Delta x = 100$, $\Delta t = 50$ (set B). All values are dimensionless.
Figure 8. One-dimensional benchmark with the continuous Galerkin (CG) scheme. Relative error $\xi_r$ as a function of the parameter $\alpha$ for different absorption functions $\sigma(x)$ and two PML thicknesses $\delta$. In each case, two sets of numerical parameters are considered: $\Delta x = 10^3$, $\Delta t = 2.5 \times 10^3$ (set A) and $\Delta x = 250$, $\Delta t = 625$ (set B). $\kappa = 1000$ for both settings. All values are dimensionless.
Figure 9. One-dimensional benchmark with the discontinuous Galerkin (DG) scheme and centered fluxes. Relative error $\xi_r$ as a function of the parameter $\alpha$ for different absorption functions $\sigma(x)$ and two PML thicknesses $\delta$. In each case, two sets of numerical parameters are considered: $\Delta x = 10^3$, $\Delta t = 2.5 \times 10^3$ (set A) and $\Delta x = 250$, $\Delta t = 625$ (set B). All values are dimensionless.
Figure 10. One-dimensional benchmark with the discontinuous Galerkin (DG) scheme and upwind fluxes. Relative error $\xi_r$ as a function of the parameter $\alpha$ for different absorption functions $\sigma(x)$ and two PML thicknesses $\delta$. In each case, two sets of numerical parameters are considered: $\Delta x = 10^{-3}$, $\Delta t = 2.5 \times 10^{-3}$ (set A) and $\Delta x = 250$, $\Delta t = 625$ (set B). The level of the dashed line corresponds to the relative error obtained with the ABC. All values are dimensionless.
close to $10^{-3}$, $10^0$ or $10^2$ (Figure 9). By contrast, for the functions $\sigma_h$ and $\sigma_{sh}$, the best $\alpha$ is always in a single range of values $10^{-1} - 10^1$.

Figure 11 shows the minimum $\xi_r$ and the corresponding optimum $\alpha$ as functions of the number of grid cells in the layer. These optimum values are obtained using the line-search algorithm. They are searched in the range $10^{-4} - 10^{-2}$ for polynomial functions, and $10^{-1} - 10^1$ for hyperbolic ones. Although there is no guarantee that the obtained values correspond to the global minimum of $\xi_r(\alpha)$, all the curves show coherent behaviors: by increasing the thickness of the layer, the minimum relative error $\xi_r$ decreases.

For each absorption function, the optimum value of $\alpha$ is similar in all discrete contexts (Figure 11, right). It lies in the range $10^{-3} - 10^{-2}$ for polynomial functions ($\sigma_2$ and $\sigma_3$), and close to $10^0$ for hyperbolic ones ($\sigma_h$ and $\sigma_{sh}$). In the first case, this optimum value varies slightly with the numerical scheme, and diminishes when the layer thickness increases. With the hyperbolic functions, it is almost always constant, except for $\sigma_{sh}$ with the DG scheme and centered fluxes, where the optimum value slightly oscillates.

The effectiveness of the different absorption functions at their best, i.e. when the optimum $\alpha$ is used, changes depending on the numerical method (Figure 11, left). With the FD scheme, the optimized polynomial functions give the smallest relative error. Among the hyperbolic functions, the shifted one $\sigma_{sh}$ works better than the other. With the CG scheme, $\sigma_h$ outperforms all the other optimized functions, which give equivalent results. Finally, with the DG scheme and centered fluxes, the functions $\sigma_2$, $\sigma_3$ and $\sigma_{sh}$ are the best, and nearly equivalent.

The performance of the hyperbolic and shifted hyperbolic functions and the optimum value of their free-parameter $\alpha$ can be interpreted by considering the particular shape of the obtained solutions. Indeed, since there is no modeling error with these functions, the global error is entirely due to the numerical error, which depends on the ability of the numerical scheme to reproduce the solution, as highlighted in Section 4. Therefore, the best absorption function corresponds to the most adequate solution shape in a given numerical context. When the functions $\sigma_h(x)$ and $\sigma_{sh}(x)$ are used with $\alpha = c$, the plane-wave solution (7) becomes respectively

$$p(x, t) = Pe^{i(k \cdot x - \omega t)} \left[ \left( 1 - \frac{x}{\delta} \right) \right] \cos \theta$$

and

$$p(x, t) = Pe^{i(k \cdot x - \omega t)} \left[ e^{x/\delta} \left( 1 - \frac{x}{\delta} \right) \right] \cos \theta.$$ \hspace{1cm} (24)

If the plane waves are normal to the interface (i.e. $\theta = 0$), their decay is then respectively linear and exponential-linear, as illustrated in Figure 12. Since linear variations are accurately represented by numerical schemes on regular grids, the decay induced by $\sigma_h$ is ideal. In particular, in a previous work [35], we showed that the linear decay of waves is perfectly captured by the FD scheme if $\alpha$ is equal to the group velocity $c_g$ (27)\textsuperscript{1}. Therefore, if the spatial step is small enough to capture the oscillations, the solution is correctly simulated. We expect the same behavior with other numerical schemes. However, at the interface between the domain $\Omega$ and the layer $\Sigma$, the spatial derivative of the plane-wave solution is discontinuous (Figure 12a). Such a solution can be correctly represented with the CG and DG schemes, whose piecewise polynomial solutions are only continuous (not continuously differentiable) and discontinuous, respectively, at the interface between elements. By contrast, the classical FD scheme considered here is based on the direct replacement of partial derivatives using finite differences in the equations. Since the partial derivative is discontinuous at the interface between $\Omega$ and $\Sigma$, this strategy is not efficient without specific interface treatment (e.g. an ad hoc transmission condition). This explains the observed results for the one-dimensional benchmark: $\sigma_h$ is good with both CG and DG schemes and bad with the FD one. The shifted hyperbolic function $\sigma_{sh}$ provides a nonlinear decay of plane waves, without discontinuity of the spatial derivative at the interface, as illustrated in Figure 12b. The shape of this solution exhibits a slow decay that is accurately represented by the FD scheme without specific interface treatment.

\textsuperscript{1}This assumption is reasonable since $c_g$ differs from the propagation velocity $c$ only for wavelengths close to the spatial step $\Delta x$ [26]. However, in practical cases, $\Delta x$ is chosen small enough in comparison to the considered wavelengths.
Figure 11. One-dimensional benchmark. Minimum relative error $\xi_r$ (left) and corresponding optimum parameter $\alpha$ (right) as a function of the PML thickness $\delta$ for the different absorption functions $\sigma(x)$. When $\delta$ increases, the spatial step $\Delta x$ remains constant and the number of cells in the PML increases. For all schemes, $\Delta x = 10^3$. The time step $\Delta t$ is 500 (FD) or $2.5 \times 10^3$ (CG and DG). All values are dimensionless.
5.2. Optimization for oblique waves

In order to test the PML with waves entering in the layer with an oblique incidence, we now consider a two-dimensional benchmark. A Gaussian-shaped pulse is initially prescribed in the center of a squared domain using the initial condition on $p(x, t)$:

$$p(x, 0) = \exp\left(-\frac{||x||^2}{R^2}\right),$$

where $R$ is a constant parameter. At the initial instant $t = 0$, the field $u(x, t)$ and the additional field $q(x, t)$, are equal to zero in $\Omega \cup \Sigma$ and $\Sigma$, respectively, and $p(x, t)$ is negligible outside $\Omega$. As time goes by, the pulse collapses and circular waves appear. The simulation stops when the main wave reaches the lateral boundary of the domain (see the snapshots of the reference solution in Figure 13). The homogeneous Dirichlet condition $u \cdot n = 0$ is prescribed at each border of the square, where $n$ is the outward normal.

The PML is tested considering a version of the problem with a truncated domain: the upper part of the squared domain is removed, and replaced with a PML $\Sigma$ (see Figure 13). The goal is to reproduce the reference solution in the truncated domain $\Omega$. The effectiveness of the PML is quantified by using the relative error $\xi_r$ that is now defined as

$$\xi_r = \sqrt{\frac{E_{\text{error with pml}}(t_f)}{E_{\text{error with wall}}(t_f)}},$$

where the global energies $E_{\text{error with pml}}(t_f)$ and $E_{\text{error with wall}}(t_f)$ are now associated to the error on the fields in the truncated domain $\Omega$. The former corresponds to the numerical solution obtained with the PML, while the latter is obtained in the case where the interface is perfectly reflective ($u \cdot n = 0$ is prescribed at the interface). In this dimensionless context, the first global energy reads

$$E_{\text{error with pml}}(t) = \int_{\Omega} \left( (p_{\text{pml}}(x, t) - p_{\text{ref}}(x, t))^2 + ||u_{\text{pml}}(x, t) - u_{\text{ref}}(x, t)||^2 \right) \, dx,$$

where the numerical fields $p_{\text{pml}}$ and $u_{\text{pml}}$ are obtained with the PML, while $p_{\text{ref}}$ and $u_{\text{ref}}$ correspond to the reference numerical solution obtained with the original squared domain (see Figure 13, left). The layer is then perfectly absorbing for $\xi_r = 0$, and perfectly reflective for $\xi_r = 1$. In the latter
**Figure 13.** Two-dimensional benchmark. Field $p(x, t)$ at different instants of the simulation in the reference domain (left) and the truncated domain $\Omega$ extended with a PML $\Sigma$ (center). The difference between these two solutions in the truncated domain is plotted on the right.

case, the energy reflected by the layer is indeed the same as with a perfectly reflective boundary condition.

The relative error $\xi_r$ is computed with numerous values of the parameter $\alpha$ for the different absorption functions and FD, CG and DG numerical schemes. Figure 14 shows the curves of the relative error $\xi_r$ as a function of $\alpha$. The spatial dimensions are $8 \times 8 \times 8$ for the reference squared domain, and $8 \times 5 \times 5$ for the truncated domain. We consider the PML thickness $\delta = 10^4$, the characteristic length of the Gaussian-shaped pulse $R = 5 \times 10^3$ and the simulation duration $t_f = 3.6 \times 10^3$. For the FD scheme, the spatial steps are $\Delta x = \Delta y = 500$ and the time-stepping leapfrog scheme is used with the time step $\Delta t = 300$. The CG and DG schemes are described in Appendix B. For both, the mesh is made of first-order triangular elements with the edge length $\ell \approx 1.75 \times 10^3$, and the Crank-Nicolson time-stepping scheme is used with the time step...
Figure 14. Two-dimensional benchmark. Relative error $\xi_r$ as a function of the parameter $\alpha$ for the different absorption functions in different numerical contexts. In Figure 14e, the level of the horizontal line indicates the relative error obtained with the ABC.
\[\Delta t = 3 \times 10^3\]. We consider two values of the stabilization parameter \(\kappa\) for the CG scheme, and both centered and upwind fluxes for the DG scheme.

As shown in Figure 14e, the PML discretized with the DG scheme and upwind fluxes is reflective for large values of \(\alpha\), by contrast with the one-dimensional case. There is now an optimum \(\alpha\) (at the minimum of each curve) for each absorption function, like with the other numerical schemes. In each case, the minimum relative error \(\xi_r\) is smaller than the one obtained with the ABC. As expected, the ABC is only approximate for oblique outgoing waves, which can be treated by the PML. The PML is then more accurate in this context, but it requires to choose a value of \(\alpha\) neither too large nor too small.

As an another difference with the one-dimensional case, the hyperbolic function \(\sigma_h\) provides now worse results than other absorption functions in all the cases. When all functions are optimized, the relative error corresponding to \(\sigma_h\) is indeed the worst in every discrete context, while this function was efficient with both CG and DG schemes in one dimension. This contrasts with Bermúdez et al. [8] who obtained better results with \(\sigma_h\) in a two-dimensional case with circular waves. This can be explained by the different numerical schemes: Bermúdez et al. used a CG scheme without stabilization based on the Helmholtz equation, which provides a different approximation of the solution. The stabilization technique employed in the present work probably plays an important role, since \(\sigma_h\) gets worse when the stabilization parameter \(\kappa\) increases (see Figures 14c–14d). By contrast, the performance of other functions is not significantly affected. Therefore, \(\sigma_h\) provides contrasted performances and is less robust than others.

The cubic function \(\sigma_3\) and the shifted hyperbolic one \(\sigma_{sh}\) exhibit similar relative error \(\xi_r\) when they are optimized. The former is slightly better than the latter for the FD scheme and the DG scheme with centered fluxes. The converse holds for the CG scheme with \(\kappa = 200\). The parabolic function \(\sigma_2\) is always worse than \(\sigma_3\).

As for the one-dimensional case, the optimum \(\alpha\) of the functions \(\sigma_h\) and \(\sigma_{sh}\) remains close to 1, while it is now systematically a bit smaller than 1 for \(\sigma_h\). For polynomial functions \(\sigma_2\) and \(\sigma_3\), the optimum \(\alpha\) is always close to \(10^{-3}\). This value is coherent with those obtained in one dimension with the same PML thickness (see Figure 11 for a PML with 10 cells), while being currently devoided of any interpretation.

### 6. CONCLUSION

PMLs provide an efficient way to simulate the truncation of a large or infinite domain. However, their main features are altered when numerical methods are used. While the perfect matching was automatic and the perfect absorption was ensured by a large absorption function \(\sigma(x)\) at the continuous level, the parameters of a discretized PML must be carefully chosen in order to limit both modeling and numerical errors.

The analytical and numerical results presented in this paper highlight the importance of the absorption function \(\sigma\), the layer thickness \(\delta\) and the spatial discretization in this issue. In a discrete context, the effectiveness of the PML critically depends on the ability of the numerical scheme to accurately represent the decay of outgoing waves in the layer. Increasing the layer thickness \(\delta\) or refining the mesh can then improve the PML, but at the prize of an increase of the computational cost. By contrast, taking a better \(\sigma\) improves the effectiveness of the layer at no additional cost. Looking for efficient absorption functions is then a very attractive approach.

Polynomial functions are widely used as absorption functions, even if they require an optimization procedure or empirical formula to choose their free parameter \(\alpha\). In this work, we confirm the effectiveness of these functions, but the choice of the parameter remains an open issue: the optimum value does not find any direct interpretation. By contrast, for the functions \(\sigma_h\) and \(\sigma_{sh}\), the optimum value for \(\alpha\) is always very close to the propagation velocity \(c\). They can therefore be used without any tuning, as already observed by Bermúdez et al. [7] and in our previous works [35, 36] in more specific contexts. We recommend the use of the shifted hyperbolic function \(\sigma_{sh}\), which is as efficient as optimized polynomial functions and more robust than \(\sigma_h\). This recommendation concerns a wide range of cases since fundamental one- and two-dimensional benchmarks are used with various
classical numerical schemes (based on finite difference, finite volume, stabilized continuous finite element and discontinuous finite element methods).

Many topics of further research about perfectly matched layers remains open. In the continuation of this paper, the knowledge of error estimates should permit a better choice of the PML thickness and the mesh refinement. Then, as mentioned in the introduction, the use of modified discrete schemes or mesh adaptation, as well as specific transmission conditions at the interface, can also improve the PML effectiveness. However, since these strategies change the approximation of the solution, the conclusion of the present work could not be relevant, and ad hoc analyses should be performed. In a general way, verifications should be made if numerical schemes not considered in this work are used (e.g. finite elements schemes with other elements and basis functions). Finally, detailed analyses and interpretation of optimum parameters should be proposed for other efficient PML formulations where other additional parameters are introduced (e.g. [15, 18, 32, 43]).

A. PLANE-WAVE ANALYSES IN THE DISCRETE FINITE DIFFERENCE CONTEXT

When a differential problem is discretized, the qualitative properties of its solution change. As dispersion and dissipation of waves are modified in the context of wave propagation, the features of PMLs are equally altered. In this appendix, the dispersion and dissipation properties of waves are studied for the basic one-dimensional problem with and without PML terms, discretized with the finite difference method. The discrete reflection coefficients associated to both infinite and finite PMLs are then derived for the constant absorption function $\bar{\sigma}$.

A.1. Discrete scalar wave system without PML terms

Let us consider the finite difference (FD) scheme for the one-dimensional scalar wave system, given by the semi-discrete system (15) without the sources terms. The dispersion and dissipation properties of waves supported by this system are studied by considering the elementary plane-wave solution (16)–(17). Injecting this solution in the system gives

$$
\begin{align*}
-\omega P + a \frac{e^{ik\Delta x/2} - e^{-ik\Delta x/2}}{\Delta x} U &= 0, \\
-\omega U + b \frac{e^{ik\Delta x/2} - e^{-ik\Delta x/2}}{\Delta x} P &= 0,
\end{align*}
$$

$$
\begin{align*}
\frac{\omega \Delta x}{2} P + a \sin \left( \frac{k \Delta x}{2} \right) U &= 0, \\
\frac{\omega \Delta x}{2} U + b \sin \left( \frac{k \Delta x}{2} \right) P &= 0.
\end{align*}
$$

(25)

The last system has non-trivial solutions if

$$
\omega = \pm \frac{2}{\Delta x} \sin \left( \frac{k \Delta x}{2} \right), \\
P = \pm \sqrt{\frac{a}{b}} U,
$$

(26)

with $c = \sqrt{ab}$. The plus sign corresponds to $k > 0$ and the minus sign is for $k < 0$. The phase velocity and the group velocity of waves are then

$$
c_p \triangleq \frac{\omega}{|k|} = \frac{2}{k \Delta x} \sin \left( \frac{k \Delta x}{2} \right), \\
c_g \triangleq \frac{\partial \omega}{\partial k} = \pm c \cos \left( \frac{k \Delta x}{2} \right).
$$

(27)

Since the phase velocity $c_p$ depends on $k$, the scheme is dispersive. Since the elementary plane-wave solution is not damped, the scheme is non-dissipative.
A.2. Discrete scalar wave system with PML terms and a constant function $\tilde{\sigma}$

Let us consider the FD scheme (15) for the one-dimensional problem with PML terms and a constant absorption function $\tilde{\sigma}$. Injecting the elementary plane-wave time-harmonic solution (19)–(20) in the system (15) gives

$$\begin{cases}
-\omega P + a \frac{e^{i\beta \Delta x/2} - e^{-i\beta \Delta x/2}}{\Delta x} U = -\bar{\sigma} P, \\
-\omega U + b \frac{e^{i\beta \Delta x/2} - e^{-i\beta \Delta x/2}}{\Delta x} P = -\bar{\sigma} U.
\end{cases}$$

$$\Leftrightarrow \begin{cases}
- (\omega + i\tilde{\sigma}) \frac{\Delta x}{2} P + a \sin \left(\frac{\beta \Delta x}{2}\right) U = 0, \\
- (\omega + i\tilde{\sigma}) \frac{\Delta x}{2} U + b \sin \left(\frac{\beta \Delta x}{2}\right) P = 0,
\end{cases}$$

This system has non-trivial solutions if

$$\omega + i\tilde{\sigma} = \pm c \frac{2}{\Delta x} \sin \left(\frac{\beta \Delta x}{2}\right), \quad \text{(28)}$$

$$P = \pm \sqrt{\frac{a}{b}} U. \quad \text{(29)}$$

From the first relation, one has an expression for the complex parameter $\beta$,

$$\beta = \pm \frac{2}{\Delta x} \arcsin \left(\frac{\Delta x \omega + i\bar{\sigma}}{c}\right).$$

Because the real part of $\beta$ non-linearly depends on the frequency, the waves are dispersive. If $\tilde{\sigma}$ is nonzero, the imaginary part of $\beta$ is positive and the waves are then damped.

Similarly to the continuous case (Section 3), the matching of the layer at the interface is studied considering a one-dimensional problem defined on the domain $\Omega = \mathbb{R}^-$ extended with the infinite PML $\Sigma = \mathbb{R}^+$. At the interface, the field $\tilde{u}_0(t)$ is governed by

$$\frac{d\tilde{u}_0}{dt} + b \frac{\tilde{P}_{1/2} - \tilde{P}_{-1/2}}{\Delta x} = -\bar{\sigma} \tilde{u}_0. \quad \text{(30)}$$

The solution of this problem is written as the superposition of incident, reflected and transmitted waves, i.e.

$$\tilde{P}_{i+1/2}(t) = \begin{cases}
P_i^i e^{i(k(i+1/2)\Delta x - \omega t)} + P_r e^{i(-k(i+1/2)\Delta x - \omega t)}, & \text{for } i = \ldots, -2, -1 \\
P_i^r e^{i(k(i+1/2)\Delta x - \omega t)}, & \text{for } i = 0, 1, \ldots
\end{cases}$$

$$\tilde{u}_i(t) = \begin{cases}
U_i^i e^{i(ki\Delta x - \omega t)} + U_r e^{i(-ki\Delta x - \omega t)}, & \text{for } i = \ldots, -1, 0 \\
U_i^r e^{i(ki\Delta x - \omega t)}, & \text{for } i = 0, 1, \ldots
\end{cases}$$

with $k > 0$. Injecting this solution in the governing equation of $\tilde{u}_0(t)$ (30) gives

$$(\bar{\sigma} - i\omega)(U^i + U^r) + \frac{b}{\Delta x} \left( P^i e^{i\beta \Delta x/2} - P^i e^{-i\beta \Delta x/2} - P^r e^{i\beta \Delta x/2} + P^r e^{-i\beta \Delta x/2} \right) = 0.$$

Using the continuity condition of the field $\tilde{u}_i$ at the interface (i.e. $U^i = U^i + U^r$) and the amplitude relations (26) and (29), it becomes

$$(\bar{\sigma} - i\omega)(P^i - P^r) + \frac{c}{\Delta x} \left( (P^i - P^r) e^{i\beta \Delta x/2} - P^i e^{-i\beta \Delta x/2} - P^r e^{i\beta \Delta x/2} \right) = 0.$$
Finally, one has the discrete reflection coefficient

\[ r_{\text{interf}}^* = \frac{P_r}{P_i} = \left| \frac{e^{-i\beta\Delta x/2} - e^{-ik\Delta x/2}}{e^{-i\beta\Delta x/2} + e^{ik\Delta x/2}} \right|. \]

The discrete reflection coefficient is now derived for a PML of finite thickness, \( i.e. \Sigma = [0, \delta] \). Let us consider an elementary solution written as the superposition of incident waves, their reflections in the domain \( \Omega \), transmitted waves in the PML and their reflections by the outer boundary of the layer \( (x = \delta) \). It reads

\[ \tilde{p}_{i+1/2}(t) = \begin{cases} P^i e^{i(k+i/2)\Delta x - \omega t)} + P^r e^{i(-k+i/2)\Delta x - \omega t)}, & \text{for } i = \ldots, -2, -1 \\ P^i e^{i(k+i/2)\Delta x - \omega t)} + P^b e^{i(-k+i/2)\Delta x - \omega t)}, & \text{for } i = 0, 1, \ldots \end{cases} \]

\[ \tilde{u}_i(t) = \begin{cases} U^i e^{i(k\Delta x - \omega t)} + U^r e^{i(-k\Delta x - \omega t)}, & \text{for } i = \ldots, -1, 0 \\ U^i e^{i(k\Delta x - \omega t)} + U^b e^{i(-k\Delta x - \omega t)}, & \text{for } i = 0, 1, \ldots \end{cases} \]

with \( k > 0 \) and \( \Re(e(\beta)) > 0 \). The continuity condition at the interface domain/layer and the boundary condition at the outer boundary of the layer \( (\tilde{u}_{N_f} = 0) \) gives the relations between the amplitudes

\[ U^i + U^r = U^t + U^b, \quad U^t e^{i\beta\delta} + U^b e^{-i\beta\delta} = 0, \]

and then

\[ U^t = \frac{U^i + U^r}{1 - e^{2i\beta\delta}}, \quad U^b = \frac{U^i + U^r}{1 - e^{-2i\beta\delta}}. \]  \( (31) \)

The governing equation of \( u_0(t) \) \( (30) \) gives

\[ (\ddot{\sigma} - i\omega)(U^i + U^r) + \frac{b}{\Delta x} \left( P^t e^{i\beta\Delta x/2} + P^b e^{-i\beta\Delta x/2} - P^i e^{-ik\Delta x/2} - P^r e^{ik\Delta x/2} \right) = 0. \]

Using the dispersion relation \( (28) \) with the plus sign and the amplitude relations \( (26) \) and \( (29) \), this relation becomes

\[ \left( e^{-i\beta\Delta x/2} + e^{i\beta\Delta x/2} \right) (U^i + U^r) + \left( U^t e^{i\beta\Delta x/2} - U^b e^{-i\beta\Delta x/2} - U^i e^{-ik\Delta x/2} + U^r e^{ik\Delta x/2} \right) = 0. \]

Using the relations \( (31) \) and the amplitude relations, one has

\[ \left( e^{-i\beta\Delta x/2} - e^{i\beta\Delta x/2} \right) \left( P^i - P^r \right) - P^i e^{-ik\Delta x/2} - P^r e^{ik\Delta x/2} = 0. \]

Finally, the discrete reflection coefficient reads

\[ r_{\text{pml}}^* = \frac{P_r}{P_i} = \left| \frac{\cos(\beta\delta + \beta\Delta x/2) - \sin(\beta\delta)}{\cos(\beta\delta + \beta\Delta x/2) + \sin(\beta\delta)} \right|. \]

Assuming the frequencies are low, \( i.e. \omega << c/\Delta x \), the formulas of the discrete reflection coefficients become

\[ r_{\text{interf}}^* \bigg|_{\omega << c/\Delta x} \approx \frac{e^\tau - 1}{e^\tau + 1}, \]

\[ r_{\text{pml}}^* \bigg|_{\omega << c/\Delta x} \approx \frac{\cosh(\tau + 2\delta\tau/\Delta x) - \sinh(2\delta\tau/\Delta x)}{\cosh(\tau + 2\delta\tau/\Delta x) + \sinh(2\delta\tau/\Delta x)}, \]

with

\[ \tau = \text{arcsinh} \left( \frac{\bar{\sigma} \Delta x}{2c} \right), \]

and are independent of the frequency.
B. NUMERICAL SCHEMES

B.1. Continuous Galerkin finite element scheme

With the CG method, the exact continuous solution is approached by a piecewise polynomial solution that verifies a weak form of the equations over the domain. This weak form is obtained by multiplying the equations (4) by test functions, integrating them over \( \Omega = \Omega \cup \Sigma \) and using integration by part. It then reads

\[
\begin{aligned}
&\left( \frac{\partial \bar{p}}{\partial t}, \bar{p} \right)_{\Omega} - \left( \mathbf{u}, \nabla \bar{p} \right)_{\Omega} = - \left( \sigma q, \bar{p} \right)_{\Omega}, \\
&\left( \frac{\partial \mathbf{u}}{\partial t}, \mathbf{u} \right)_{\Omega} - \left( \nabla \mathbf{u}, \bar{u} \right)_{\Omega} = - \left( \sigma (\mathbf{e}_x \cdot \mathbf{u}) \mathbf{e}_x, \bar{u} \right)_{\Omega}, \\
&\left( \frac{\partial q}{\partial t}, \bar{q} \right)_{\Sigma} - \left( (\mathbf{e}_x \cdot \mathbf{u}) \mathbf{e}_x, \bar{q} \right)_{\partial \Omega} + \left( (\mathbf{e}_x \cdot \mathbf{u}) \mathbf{e}_x, \nabla \bar{q} \right)_{\Sigma} = - \left( \sigma q, \bar{q} \right)_{\Sigma},
\end{aligned}
\]  

(32)

where we use the notations \((i,j)_A = \int_A (i \cdot j) \, dA\), \((i,j)_n = \int_{\partial A} (i \cdot j) \cdot n \, dB\) and \((i,j)_{\partial A} = \int_{\partial A} (i \cdot j) \cdot n \, dB\) for the integrals over the region \(A\) and its boundary, and \(\bar{p}, \mathbf{u}\) and \(\bar{q}\) are the test functions. The vector \(n\) is the outward normal on the boundary \(\partial A\). However, this formulation generates spurious numerical oscillations, even without the PML terms and the additional equation. Stabilization techniques have been developed to avoid them (see e.g. [24]). In particular, the Pressure-Stabilization Petrov-Galerkin (PSPG) method proposed by Hughes et al. [30], consists in adding a stabilisation term in the left-hand side of the continuity equation (the first equation of the system (32))

\[
S_p = - (\kappa \mathbf{R}_u, \nabla \bar{p})_{\Omega},
\]

where \(\kappa\) is the numerical stabilization parameter and \(\mathbf{R}_u\) is the residual associated to the momentum equation, i.e.

\[
\mathbf{R}_u = \frac{\partial \mathbf{u}}{\partial t} - \nabla p + \sigma (\mathbf{e}_x \cdot \mathbf{u}) \mathbf{e}_x.
\]

Considering the similarities between the first and last equations of the formulation, we add a corresponding stabilization term in the governing equation of \(q\), i.e.

\[
S_q = - (\kappa (\mathbf{e}_x \cdot \mathbf{R}_u) \mathbf{e}_x, \nabla \bar{q})_{\Sigma}.
\]

An impermeability boundary condition \(\mathbf{u} \cdot n = 0\) on \(\partial \Omega\) is weakly incorporated in the formulation by removing the boundary term of the first and last equations.

B.2. Discontinuous Galerkin finite element scheme

With the DG method, the exact continuous solution is approached by a piecewise polynomial solution, discontinuous at the interfaces, that verifies weakly the equations over each element [27]. Multiplying the equations (4) by test functions, integrating them over each element \(\Omega_e\) and using integration by part, one gets the weak form:

\[
\begin{aligned}
&\left( \frac{\partial \bar{p}}{\partial t}, \bar{p} \right)_{\Omega_e} + (\mathbf{u}^*, \bar{p})_{\partial \Omega_e} - (\mathbf{u}, \nabla \bar{p})_{\Omega_e} = - (\sigma q, \bar{p})_{\Omega_e}, \\
&\left( \frac{\partial \mathbf{u}}{\partial t}, \mathbf{u} \right)_{\Omega_e} + (bp^*, \bar{u})_{\partial \Omega_e} - (bp, \nabla \bar{u})_{\Omega_e} = - (\sigma (\mathbf{e}_x \cdot \mathbf{u}) \mathbf{e}_x, \bar{u})_{\Omega_e}, \\
&\left( \frac{\partial \bar{q}}{\partial t}, \bar{q} \right)_{\Omega_e} + (a(\mathbf{e}_x \cdot \mathbf{u})^* \mathbf{e}_x, \bar{q})_{\partial \Omega_e} - (a(\mathbf{e}_x \cdot \mathbf{u}) \mathbf{e}_x, \nabla \bar{q})_{\Omega_e} = - (\sigma q, \bar{q})_{\Omega_e}.
\end{aligned}
\]

Because the numerical solution is discontinuous at the interface between two elements, numerical fluxes must be chosen at this interface. For a linear wave problem, the so-called upwing fluxes...
are classically defined using a Riemann solver. They are computed using the Riemann invariants $p^* = \bar{p} + n \cdot [u]$ and $u^* = \bar{u} + n[p]$, where the mean and the jump of $p$ and $u$ are defined as

$$\bar{p} = \frac{p^+ + p^-}{2}, \quad [p]^* = \frac{p^+ - p^-}{2}, \quad \bar{u} = \frac{u^+ + u^-}{2}, \quad [u]^* = \frac{u^+ - u^-}{2}. \quad (33)$$

As an alternative, the centered fluxes based on the mean values $p^* = \bar{p}$ and $u^* = \bar{u}$ can be used.

REFERENCES


