default A Perfectly Matched Layer for the Helmholtz Equation in a Semi-infinite Strip

I. Singer [a] E. Turkel [a,*]

[a] *Department of Mathematics*
Tel Aviv University
Tel Aviv, Israel

history

Abstract

The Perfectly Matched Layer (PML) has become a widespread technique for preventing reflections from far field boundaries for wave propagation problems in both the time dependent and frequency domains. We develop a discretization to solve the Helmholtz equation in an infinite two dimensional strip. We solve the interior equation using high-order finite differences schemes. The combined Helmholtz-PML problem is then analyzed for the parameters that give the best performance. We show that the use of local high-order methods in the physical domain coupled with a specific second order approximation in the PML yields global high-order accuracy in the physical domain. We discuss the impact of the parameters on the effectiveness of the PML. Numerical results are presented to support the analysis.

keywords. *Key words:* Helmholtz, PML, finite difference
1991 MSC: 65N06

1 Introduction

The Helmholtz equation

$$\Delta u + k^2 u = 0, \tag{1}$$

describes a wide variety of wave propagation phenomena including electromagnetic waves and acoustics. To solve this equation in an unbounded domain on a computer, one approach is to truncate the unbounded domain and introduce a boundary condition on the artificial outer surface. For many years the

[*] Corresponding Author.
Email address: turkel@post.tau.ac.il (E. Turkel).

standard boundary condition was a local absorbing condition that was a generalization of the Sommerfeld radiation condition e.g. [4], but in recent years a number of models based on the PML (Perfectly Matched Layers) scheme have become popular. These layers minimize the reflections caused by the artificial boundary.

Berenger [5],[6] was the first to introduce a PML for the time dependent Maxwell equations. Abarbanel and Gottlieb [1] proved that this approach is not well posed and since then several other approaches generalizing the ideas of Berenger have been suggested. A survey of PML layers is to be found in [8]. Turkel and Yefet [16] showed that several of these approaches are linearly equivalent. The solvability and the uniqueness of the PML equation for the Helmholtz equation was analyzed by Turkel and Tsynkov in [13]. In their paper, the decaying function inside the PML was assumed to be, for convenience of the analysis, a constant. This choice is not suitable for numerical computations.

We are interested in obtaining high accuracy for the approximation to the Helmholtz equation. In particular we shall consider both fourth and sixth order accurate approximations in the physical domain. Hence, we shall use a PML in the far field to minimize reflections and so hopefully maintain the high accuracy of the interior scheme. We extend [13] and search for a practical set of parameters in the PML layer based on an analysis of the error in the combined problem. We also analyze the solvability and accuracy of the PML schemes for this problem. The physical PML-Helmholtz scheme, which we develop, is then solved by using high order finite differences schemes specially designed for the Helmholtz equation. For a constant value of k in (1) we developed a sixth order accurate scheme (and a fourth order accurate scheme for a variable k [9]). We analyze the effect of the use of these schemes on the solution and we find conditions required for convergence. We also verify the assumption made in [13] that the overall accuracy of the Helmholtz-PML scheme depends only of the order of accuracy inside the physical domain where we solve the pure Helmholtz equation. We support our analysis with numerical results.

In order to invert the linear system one frequently uses iterative solvers. In the last section we construct a preconditioner specially tailored for the combined problem. This preconditioner can be used with any Krylov space method.

In designing and analyzing the PML-Helmholtz equation, we use the two dimensional notation of acoustics [13] and [16] for a waveguide. In two space dimensions this is equivalent to the TE version of Maxwell's equations. Denoting u as the pressure, we get (see for example [16])

$$\frac{\partial}{\partial x}\left(\frac{S_y}{S_x}u_x\right)+\frac{\partial}{\partial y}\left(\frac{S_x}{S_y}u_y\right)+k^2 S_x S_y u=0. \tag{2}$$

where

$$S_x = 1 + \frac{\sigma_x}{ik},\ S_y = 1 + \frac{\sigma_y}{ik}$$

and σ_x and σ_y are functions of only x, y respectively. When $\sigma_x = \sigma_y = 0$ this reduces to the Helmholtz equation (1). In the rest of the paper we only consider the strip so $\sigma_y = 0$ and hence $S_y = 1$.

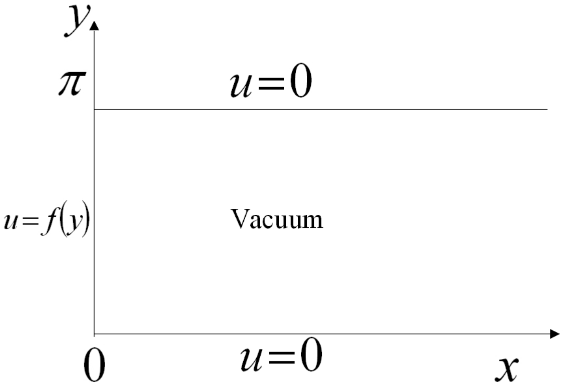

Fig. 1. The x-aligned semi infinite waveguide

2 The infinite strip problem

We consider a semi-infinite x-aligned waveguide from $x = 0$ to $x = \infty$ with width π in the y direction, see Figure 1. We wish to solve the Helmholtz equation (1) in this semi-infinite strip with the boundary conditions:

$$u\,(x, 0) = u\,(x, \pi) = 0,$$

and

$$u\,(0, y) = f\,(y)\,,\ 0 \leq y \leq \pi \quad , \qquad f\,(0) = f\,(\pi) = 0$$

and specify that u is outgoing at $+\infty$. Using the sine Fourier series expansion in the y direction we get

$$u\,(x, y) = \sum_{n=1}^{\infty} \frac{2}{\pi} \left(\int_0^{\pi} f\,(y) \sin\,(ny)\, dy \right) e^{-i\sqrt{k^2-n^2}x} \sin ny.$$

To simplify the analysis of the PML layer we assume the boundary condition at the entrance to the waveguide is

$$f\,(y) = \sin\,(my) \tag{3}$$

where m is an integer. Then we get the exact solution which we label u_{exact}

$$u_{exact}\left(x,y\right)=e^{-i\sqrt{k^2-m^2}x}\sin my. \tag{4}$$

In the case $m>k$ (evanescent wave) we get

$$u_{exact}\left(x,y\right)=e^{-\sqrt{m^2-k^2}x}\sin my$$

3 Constructing the PML

Since we cannot solve the semi-infinite problem on a computer we instead solve in a bounded domain: $[0,L_1]\times[0,\pi]$, and construct a PML. We first analyze an infinite PML and then a PML with width L_2-L_1, see Figure 2.

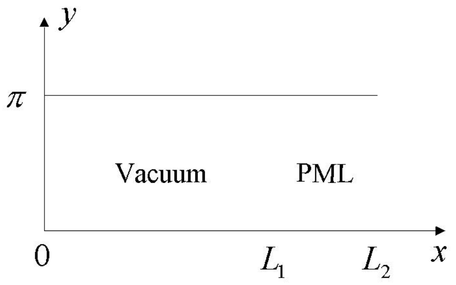

Fig. 2. The PML layer in the strip

In the interior of the strip (vacuum) we solve the Helmholtz equation and in the PML we solve equation (2). In this case we choose $\sigma_y=0$ in the PML and so $S_y\!=\!1$ and (2) becomes

$$\frac{\partial}{\partial x}\left(\frac{1}{S_x}u_x\right)+\frac{\partial}{\partial y}\left(S_xu_y\right)+k^2S_xu=0,\qquad S_x=S_x(x). \tag{5}$$

Using the Fourier series expansion in y we get

$$\frac{d}{dx}\left(\frac{1}{S_x}\frac{d\hat{u}\left(x,n\right)}{dx}\right)+S_x\left(k^2-n^2\right)\hat{u}\left(x,n\right)=0. \tag{6}$$

We set

$$S_x\left(x\right)=\begin{cases}1 & 0\leq x\leq L_1\\ 1+\frac{\sigma_x}{ik} & x>L_1\end{cases} \tag{7}$$

We generalize this by considering $S_x(x) = A + \frac{\sigma_x}{B+ik}$. Our computational results show that the error deteriorates when A differs appreciably from 1. Choosing B non-zero only slightly improves the accuracy. Equation (6) represents $\hat{u}$ in the entire space (vacuum + PML). Substituting

$$t = \int_0^x S_x(r)\,dr$$

and labeling

$$Y_n(t) = \hat{u}(x,n)$$

we get for $Y_n(t)$ an equation with constant coefficients

$$\frac{d^2}{dt^2} Y_n(t) + \left(k^2 - n^2\right) Y_n(t) = 0.$$

The solution is

$$\hat{u}(x,n) = c_+ e^{i\sqrt{k^2-n^2}\int_0^x S_x(r)dr} + c_- e^{-i\sqrt{k^2-n^2}\int_0^x S_x(r)dr} \tag{8}$$

where c_+ and c_- are arbitrary constants. The solution for (6) in an infinite PML is found when we add the condition that only the right-traveling waves are present in addition to the condition in $x = 0$.

Using (3) we get the solution for (5) which we label u_{I-pml} (Infinite-PML):

$$u_{I-pml}(x,y) = \begin{cases} e^{-i\sqrt{k^2-m^2}\int_0^x S_x(r)dr} \sin my & \text{if } m < k \\ \\ e^{-\sqrt{m^2-k^2}\int_0^x S_x(r)dr} \sin my & \text{if } m > k \end{cases}. \tag{9}$$

Comparing (4) with (9) we get for $0 \leq x \leq L_1$:

$$u_{I-pml}(x,y) = u_{exact}(x,y)$$

which shows that the PML is perfectly non-reflecting.

When solving the problem on a computer we need to truncate the semi-infinite domain at L_2. We choose the boundary condition as $u = 0$ at $x = L_2$. Choosing other types of boundary conditions at $x = L_2$ does not significantly change the solution. Hence, instead of $c_+ = 0$ in (8) we need to solve (6) with the boundary condition $\hat{u}(L_2, n) = 0$. For our choice of $f(y) = \sin(my)$, with $n \neq m$, $\hat{u}(x,n) = 0$ and boundary conditions

$$\hat{u}(0,m) = 1 \qquad \hat{u}(L_2,m) = 0.$$

we get

$$\hat{u}(x,m) = c_+ e^{i\sqrt{k^2-m^2}\int_0^x S_x(r)dr} + c_- e^{-i\sqrt{k^2-m^2}\int_0^x S_x(r)dr}$$

Solving for c_+ and c_-

$$c_+ = \frac{-1}{\eta - 1} \qquad c_- = \frac{\eta}{\eta - 1}$$

where

$$\eta = e^{2i\sqrt{k^2-m^2}\int_0^{L_2} S_x(r)dr}. \tag{10}$$

Uniqueness is not guaranteed when $\eta = 1$. Otherwise

$$2i\sqrt{k^2 - m^2}\int_0^{L_2} S_x\left(r\right)dr = 2i\sqrt{k^2 - m^2}\left(L_2 + \frac{1}{ik}\int_{L_1}^{L_2} \sigma_x\left(r\right)dr\right)$$

since $\sigma_x\left(x\right) > 0$ for $x \in (L_1, L_2]$, then for $k \neq m$ it follows that $Re\left(2i\sqrt{k^2 - m^2}\int_0^{L_2} S_x\left(r\right)dr\right)$ 0 and $\eta \neq 1$ and uniqueness follows. Henceforth, we assume that $k \neq m$.

Labeling this unique solution u_{F-pml} (Finite PML) we find

$$u_{F-pml} = \left(\frac{-1}{\eta - 1}e^{i\sqrt{k^2-m^2}\int_0^x S_x(r)dr} + \frac{\eta}{\eta - 1}e^{-i\sqrt{k^2-m^2}\int_0^x S_x(r)dr}\right)\sin my \tag{11}$$

and for the case $m < k$ (travelling)

$$\begin{aligned} error(x, y) &= u_{I-pml} - u_{F-pml} \\ &= \frac{1}{\eta - 1}\left(e^{i\sqrt{k^2-m^2}\int_0^x S_x(r)dr} - e^{-i\sqrt{k^2-m^2}\int_0^x S_x(r)dr}\right)\sin\left(my\right) \\ &= \frac{2i}{\eta - 1}\sin\left(\sqrt{k^2 - m^2}\int_0^x S_x\left(r\right)dr\right)\cdot\sin\left(my\right). \end{aligned}$$

We are interested in the solution in the interior and so we examine the error when $0 \leq x \leq L_1$

$$\begin{aligned} &\left\|error(x, y)\right\|_\infty \qquad (12) \\ &= \max_{0\leq x\leq L_1,\ 0\leq y\leq\pi}\left|\frac{2i}{\eta - 1}\sin\left(\sqrt{k^2 - m^2}x\right)\cdot\sin\left(my\right)\right| \\ &= \left|\frac{2}{\eta - 1}\right|. \end{aligned}$$

Our goal is to minimize the error and so we choose a suitable function $\sigma_x\left(x\right)$ that will minimize the value of $\left|\frac{2}{\eta-1}\right|$. We wish η in (10) to satisfy $|\eta| >> 1$. For the case $m > k$ (evanescent) we get

$$\begin{aligned} error(x,y) &= u_{I-pml} - u_{F-pml} \\ &= \frac{\eta}{\eta-1}\left(e^{-\sqrt{m^2-k^2}\int_0^x S_x(r)dr} - e^{\sqrt{m^2-k^2}\int_0^x S_x(r)dr}\right)\sin\left(my\right) \\ &= -\frac{2\eta}{\eta-1}\sinh\left(\sqrt{m^2-k^2}\int_0^x S_x\left(r\right)dr\right)\cdot\sin\left(my\right) \end{aligned}$$

and in the interior

$$\left\|error(x,y)\right\|_\infty \simeq \left|\frac{\eta}{\eta-1}e^{\sqrt{m^2-k^2}L_1}\right|. \tag{13}$$

In this case we wish in (10) that $|\eta| << 1$.

4 Minimizing the error

The function $\sigma_x : [L_1, L_2] \to \mathbf{R}$ is chosen so that it has the following properties:

$$\begin{aligned} \sigma_x\left(x\right) &> 0 \qquad \text{for } x \in \left(L_1, L_2\right] \\ \sigma_x\left(L_1\right) &= 0 \\ &\sigma_x\left(x\right) \ \text{ is smooth in } \left[L_1, L_2\right]. \end{aligned}$$

A suitable choice for this function is

$$\sigma_x\left(x\right) = \sigma\left(\frac{x-L_1}{L_2-L_1}\right)^p \tag{14}$$

where σ is a positive constant and $p \geq 1$. When evanescent waves are present it may be more advantageous to use an exponential fit for σ_x. In this study we shall only consider the polynomial fit. Integrating we get

$$\int_{L_1}^x \sigma_x\left(r\right)dr = \frac{\sigma\left(L_2-L_1\right)}{p+1}\left(\frac{x-L_1}{L_2-L_1}\right)^{p+1}$$

and

$$u_{F-pml} = \sin my\,\cdot \tag{15}$$
$$\begin{cases} \frac{-1}{\eta-1}e^{i\sqrt{k^2-m^2}x} + \frac{\eta}{\eta-1}e^{-i\sqrt{k^2-m^2}x} & 0 < x \leq L_1 \\ \frac{-1}{\eta-1}e^{i\sqrt{k^2-m^2}L_1+\sqrt{1-\varepsilon}\frac{\sigma(L_2-L_1)}{p+1}\left(\frac{x-L_1}{L_2-L_1}\right)^{p+1}} + \\ \frac{\eta}{\eta-1}e^{-i\sqrt{k^2-m^2}L_1-\sqrt{1-\varepsilon}\frac{\sigma(L_2-L_1)}{p+1}\left(\frac{x-L_1}{L_2-L_1}\right)^{p+1}} & L_1 < x \leq L_2 \end{cases}$$

where $\varepsilon = \left(\frac{m}{k}\right)^2$ and

$$\eta = e^{2i\sqrt{k^2-m^2}\int_0^{L_2} S_x(r)dr} = e^{2i\sqrt{k^2-m^2}L_2 + \frac{2\sigma\sqrt{1-\varepsilon}(L_2-L_1)}{p+1}}. \tag{16}$$

For $m < k$ we get

$$|\eta| = e^{\frac{2\sigma\sqrt{1-\varepsilon}(L_2-L_1)}{p+1}} >> 1$$

and for $m > k$

$$|\eta| = e^{-2\sqrt{m^2-k^2}L_2} << 1,$$

which are the desired results for (12) and (13). Using the estimate (12) we get for $m < k$

$$\|error(x,y)\|_\infty = \left|\frac{2}{\eta - 1}\right| \simeq \left|\frac{2}{\eta}\right| = 2e^{-\frac{2\sigma\sqrt{1-\varepsilon}(L_2-L_1)}{p+1}} \tag{17}$$

and for $m > k$ using (13) we get

$$\|error(x,y)\|_\infty \simeq \left|\frac{\eta}{\eta - 1}e^{\sqrt{m^2-k^2}L_1}\right| \simeq \left|\eta e^{\sqrt{m^2-k^2}L_1}\right| = e^{-\sqrt{m^2-k^2}(2L_2-L_1)}.$$

We see that for evanescent waves, $m > k$, the norm of the error does not depend on σ_x. To decrease the error we can only increase the length of the PML region, i.e. $2L_2 - L_1$. When $m < k$ the error depends exponentially on the value of $\frac{2\sigma(L_2-L_1)}{p+1}$. From this continuous analysis we conclude that the parameters should be chosen to satisfy the following criteria. A large value of σ, a wide PML (extend $L_2 - L_1$) and $p = 1$. Further analysis (below) for the numerical algorithm demonstrates that one should be more careful when choosing these parameters especially p.

We need a scheme to approximate the u_{F-pml} solution. We use high-order finite differences schemes in the interior. To reduce the size of the matrices we wish to minimize the number of points in the artificial perfectly matched layer. In the next sections we present these schemes and analyze their influence on the error. We will also see that the choice of a large value of σ and a small value of p is inaccurate.

5 Finite differences

Let $\phi_{i,j}$ be the numerical approximation to the $u_{F-pml}(x_i, y_j)$ solution. We wish to have a symmetric stencil in both directions x and y. A scheme having this property has the form

$$A_0\phi_{i,j} + A_s\sigma_s + A_c\sigma_c = 0$$

where

$$\sigma_s = \phi_{i,j+1} + \phi_{i+1,j} + \phi_{i,j-1} + \phi_{i-1,j}$$

is the sum of the values of the mid-side points and

$$\sigma_c = \phi_{i+1,j+1} + \phi_{i+1,j-1} + \phi_{i-1,j-1} + \phi_{i-1,j+1}$$

is the sum of the values at the corner points.

In the PML we need to solve a variable coefficient problem

$$\frac{\partial}{\partial x}\left(Au_x\right) + \frac{\partial}{\partial y}\left(Bu_y\right) + k^2 Cu = 0. \tag{18}$$

In our strip problem

$$A = \frac{1}{S_x} \qquad B = C = S_x$$

We have not found a compact formula which keeps the self-adjoint form of the equation (18) and is also more than second order accurate for non-constant A, B, C. Instead, we use high-order accurate self-adjoint schemes in the interior and automatically switch to a second order accurate scheme in the PML layer while preserving the self-adjoint property. Since the PML is artificial we are only interested in preserving the global high accuracy in the interior domain. Using a general second order accurate scheme in the PML usually corrupts the high order accuracy used in the interior, and yields an overall low accuracy. Thus, matching the schemes between the interior and the PML is very important.

We start with the standard second order three point symmetric approximation

$$D_x\left(Au_x\right)_j = \frac{A_{i+\frac{1}{2},j}\left(u_{i+1,j} - u_{i,j}\right) - A_{i-\frac{1}{2},j}\left(u_{i,j} - u_{i-1,j}\right)}{h^2}. \tag{19}$$

We construct a more general divergence free form by averaging this approximation in the j direction. So we take $\left[Au_x\right]_x = \alpha D_x\left(Au_x\right)_j + \frac{1-\alpha}{2}(D_x\left(Au_x\right)_{j+1} + D_x\left(Au_x\right)_{j-1})$ and a similar formula in the y direction. The approximation to Cu is a general nine point formula

$$\begin{aligned}
[Cu] = {} & \left(1 - 4\beta_s - 4\beta_c\right) C_{i,j} u_{i,j} \\
& + \beta_s \begin{pmatrix} C_{i+1,j} u_{i+1,j} + C_{i-1,j} u_{i-1,j} + \\ C_{i,j+1} u_{i,j+1} + C_{i,j-1} u_{i,j-1} \end{pmatrix} \\
& + \beta_c \begin{pmatrix} C_{i+1,j+1} u_{i+1,j+1} + C_{i-1,j+1} u_{i-1,j+1} + \\ C_{i+1,j-1} u_{i+1,j-1} + C_{i-1,j-1} u_{i-1,j-1} \end{pmatrix}
\end{aligned} \tag{20}$$

Thus, for $A\!=\!B\!=\!C\!=\!1$ we get

$$\begin{aligned} A_0 &= -4\alpha + (1-4\beta_s-4\beta_c)\,(kh)^2 \\ A_s &= 2\alpha - 1 + \beta_s\,(kh)^2\,, \qquad A_c = 1-\alpha+\beta_c\,(kh)^2 \end{aligned} \tag{21}$$

where h is the grid-size of the stencil. This approximation is guaranteed to be $O\left(h^2\right)$ for all values of α, β_s, β_c. Choosing $\alpha = 1, \beta_s = 0, \beta_c = 0$ recovers the standard pointwise representation which is second order accurate with

$$A_0 = -4 + (kh)^2\,, \quad A_s = 1, \quad A_c = 0. \tag{22}$$

We use higher-order schemes for the pure Helmholtz equation. Choosing

$$\alpha = \frac{5}{6}, \quad \beta_s = \frac{1}{12} - \frac{\gamma}{72}, \quad \beta_c = \frac{\gamma}{144}$$

for an arbitrary constant γ we achieve an $O\left(h^4\right)$ scheme yielding

$$\begin{aligned} A_0 &= -\frac{10}{3} + (kh)^2\left(\frac{2}{3}+\frac{\gamma}{36}\right) \\ A_s &= \frac{2}{3} + (kh)^2\left(\frac{1}{12}-\frac{\gamma}{72}\right), \quad A_c = \frac{1}{6} + (kh)^2\,\frac{\gamma}{144}. \end{aligned} \tag{23}$$

This stencil is fourth-order accurate also for variable $k\,(x,y)$ as proved in [9].Choosing $\gamma = \frac{14}{5}$ and adding $O((kh)^4)$ terms to the coefficients one can achieve sixth order accuracy [12]. In particular, for arbitrary δ, choosing

$$\alpha = \frac{5}{6}, \qquad \beta_s = \frac{2}{45} + \frac{3-2\delta}{720}\,(kh)^2\,, \quad \beta_c = \frac{7}{360} + \frac{\delta}{720}\,(kh)^2$$

$$\begin{aligned} A_0 &= -\frac{10}{3} + \frac{67}{90}\,(kh)^2 + \frac{\delta-3}{180}\,(kh)^4 \\ A_s &= \frac{2}{3} + \frac{2}{45}\,(kh)^2 + \frac{3-2\delta}{720}\,(kh)^4 \\ A_c &= \frac{1}{6} + \frac{7}{360}\,(kh)^2 + \frac{\delta}{720}\,(kh)^4\,. \end{aligned} \tag{24}$$

yields a $O\left(h^6\right)$ scheme for constant k.

Hence, when we use one of the schemes (22, 23 or 24) for the combined problem, the scheme is locally high order accurate in the interior and is second order accurate in the PML layer. The accuracy in the PML is physically irrelevant. Hence, we wish that the use of the low-order accuracy in the PML will not destroy the global high-order accuracy in the interior.

6 Exploring the numerical error

Using a Taylor expansion in the approximation (19) we get

$$D_x\left(Au_x\right)_j = \frac{\partial}{\partial x}\left(Au_x\right) + \frac{h^2}{24}\left(\frac{\partial^3}{\partial x^3}\left(Au_x\right) + \frac{\partial}{\partial x}\left(Au_{xxx}\right)\right) + O\left(h^4\right).$$

Averaging in the y direction yields

$$\begin{aligned}\left[Au_x\right]_x &= \alpha D_x\left(Au_x\right)_j + \frac{1-\alpha}{2}\left(D_x\left(Au_x\right)_{j+1} + D_x\left(Au_x\right)_{j-1}\right)\\ &= \frac{\partial}{\partial x}\left(Au_x\right) + \nu h^2 + O\left(h^4\right)\end{aligned}$$

where

$$\nu = \frac{1}{24}\left(\frac{\partial^3}{\partial x^3}\left(Au_x\right) + \frac{\partial}{\partial x}\left(Au_{xxx}\right) + 12\left(1-\alpha\right)\frac{\partial^2}{\partial y^2}\frac{\partial}{\partial x}\left(Au_x\right)\right).$$

Using the Taylor expansion we find that approximation (20) satisfies

$$[Cu] = Cu + h^2\left(\beta_s + 2\beta_c\right)\left(\left(Cu\right)_{xx} + \left(Cu\right)_{yy}\right) + O\left(h^4\right).$$

Therefore, the finite difference formula is equivalent to

$$\frac{\partial}{\partial x}\left(\frac{u_x}{S_x}\right) + \frac{\partial}{\partial y}\left(S_x u_y\right) + k^2 S_x u + \Theta_{h^2} h^2 + O(h^4), \tag{25}$$

where

$$\begin{aligned}\Theta_{h^2} = &\tfrac{1}{24}\left(\tfrac{\partial^3}{\partial x^3}\left(\tfrac{u_x}{S_x}\right) + \tfrac{\partial}{\partial x}\left(\tfrac{u_{xxx}}{S_x}\right) + 2S_x u_{yyyy}\right) +\\ &\tfrac{1-\alpha}{2}\left(\tfrac{\partial^2}{\partial y^2}\tfrac{\partial}{\partial x}\left(\tfrac{u_x}{S_x}\right) + \tfrac{\partial^2}{\partial x^2}\left(S_x u_{yy}\right)\right) +\\ &k^2\left(\beta_s + 2\beta_c\right)\left(\left(S_x u\right)_{xx} + S_x u_{yy}\right).\end{aligned} \tag{26}$$

In order for the approximation to be accurate we require that $h^2\Theta_{h^2} << 1$.

Using $u = u_{F-pml}$ in (26), assuming for the analysis $m << k$ and $k >> 1$, and collecting the $O\left(k^4\right)$ terms we get

$$\Theta_{h^2} \simeq c_{\pm}\left(k^2 - m^2\right)^2\left(\frac{1}{12} - \mu\right) S_x^3 e^{\pm\sqrt{k^2-m^2}\int_0^x S_x(r)dr} \tag{27}$$

where

$$\mu = \frac{\left(\beta_s + 2\beta_c\right)}{1-\varepsilon}, \quad \varepsilon = \left(\frac{m}{k}\right)^2.$$

Using the choice of σ_x (14), the exact values of the constants $c_\pm$ in the solution u_{F-pml} (15) and denoting

$$z = \frac{x - L_1}{L_2 - L_1},\ 0 \leq z \leq 1, \tag{28}$$

we get inside the PML the approximation

$$|\Theta_{h^2}| \simeq \left(k^2 - m^2\right)^2 \left(\frac{1}{12} - \mu\right) e^{-\sqrt{1-\varepsilon}\frac{\sigma(L_2-L_1)}{p+1} z^{p+1}} \left(1 + \left(\frac{\sigma}{k}\right)^2 z^{2p}\right)^{\frac{3}{2}}.$$

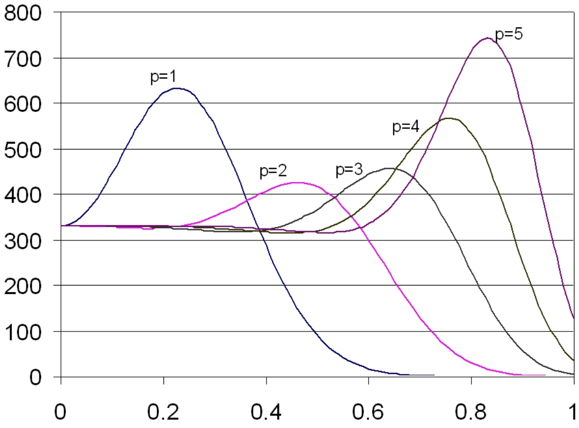

Fig. 3. Value of $|\Theta_{h^2}|$ for $k = 8,\ m = 1,\ \sigma = 50,\ L_2 - L_1 = \frac{\pi}{4}$

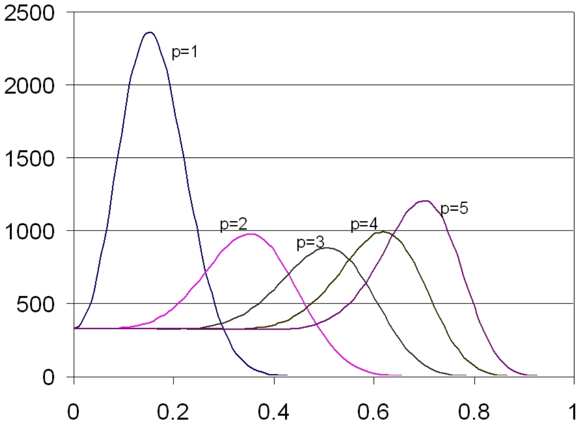

Fig. 4. Value of $|\Theta_{h^2}|$ for $k = 8,\ m = 1,\ \sigma = 150,\ L_2 - L_1 = \frac{\pi}{4}$

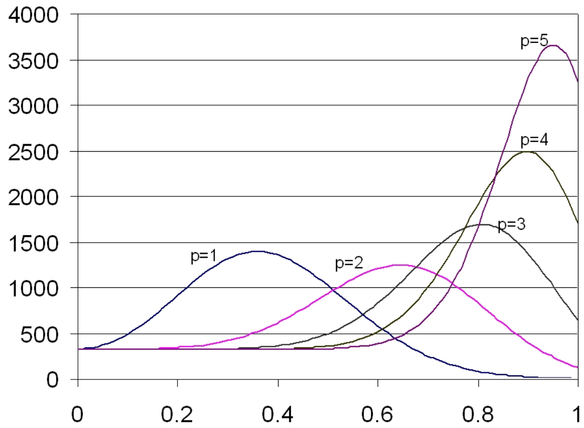

Fig. 5. Value of $|\Theta_{h^2}|$ for $k = 8,\ m = 1,\ \sigma = 50,\ L_2 - L_1 = \frac{\pi}{8}$

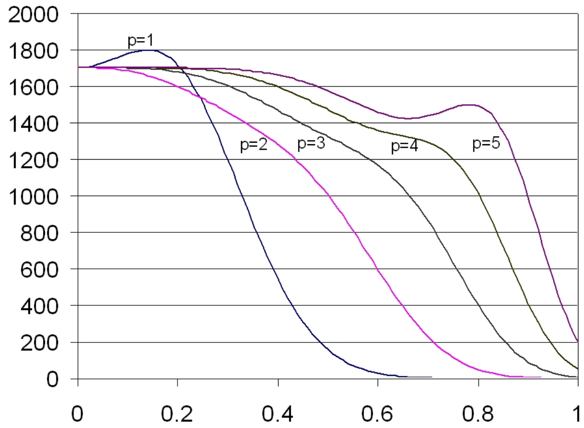

Fig. 6. Value of $|\Theta_{h^2}|$ for $k = 12,\ m = 1,\ \sigma = 50,\ L_2 - L_1 = \frac{\pi}{4}$

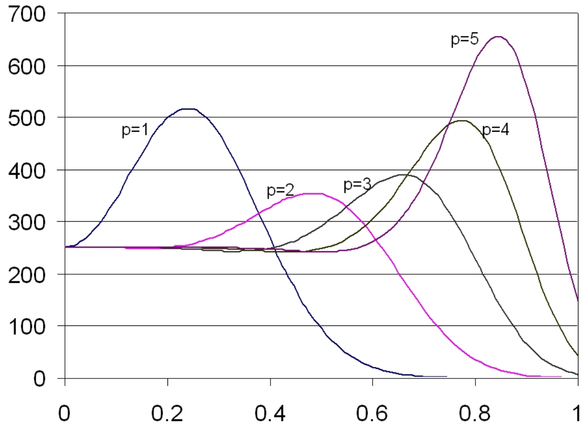

Fig. 7. Value of $|\Theta_{h^2}|$ for $k = 8,\ m = 3,\ \sigma = 50,\ L_2 - L_1 = \frac{\pi}{4}$

We see in figures 3-7 the behavior of $|\Theta_{h^2}|$ for various values of p. In these figures the x-axis of the graph is the variable z, defined in (28). We set $\beta_s + 2\beta_c = 0$ which is what is used in the standard pointwise representation (22). In most cases p should be set as 2 or 4. Increasing the value of σ increases also the value of $|\Theta_{h^2}|$. These two facts contradict our desire to decrease the error (17). Increasing $L_2 - L_1$ decreases the value of $|\Theta_{h^2}|$ as well as the error, but increases the work. A thick PML requires more storage and CPU and also means a harder task for the linear solvers.

An interesting result is found when examining the value of $|\Theta_{h^2}|$ near the intersection between the vacuum and the PML layer, i.e. $z \to 0^+$. At this point we can find a lower bound of $|\Theta_{h^2}|$ and get

$$|\Theta_{h^2}| \simeq \left(k^2 - m^2\right)^2 \left(\frac{1}{12} - \mu\right). \tag{29}$$

Combining this result with Figures 6 and 7 we conclude that as $(k^2 - m^2)^2$ increases, the schemes become more inaccurate independent of the values of the other parameters, σ, p and $L_2 - L_1$. This requires one to refine the grid.

We can also decrease the value of $|\Theta_{h^2}|$ by choosing $\mu \to \frac{1}{12}$. When we examine the value of μ in the high-order schemes (23, 24) we get

$$\frac{1}{12} - \mu = \frac{1}{12}\frac{m^2}{k^2 - m^2}.$$

Hence, for high order accurate schemes, (27) is not valid and we get instead

$$|\Theta_{h^2}| \simeq \frac{\sigma p}{12\left(L_2 - L_1\right)}\sqrt{1-\varepsilon}\left(k^2 - 4m^2\right) e^{-\sqrt{1-\varepsilon}\frac{\sigma(L_2-L_1)}{p+1}z^{p+1}} z^{p-1}\sqrt{1 + \left(\frac{\sigma}{k}\right)^2 z^{2p}}. \tag{30}$$

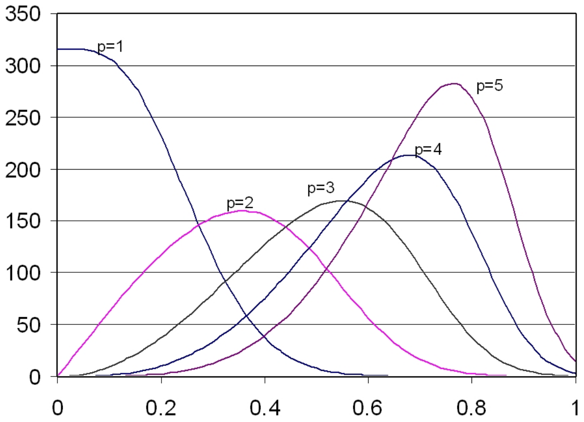

Fig. 8. Value of $|\Theta_{h^2}|$ in the high-order schemes for $k = 8,\ m = 1,\ \sigma = 50,\ L_2 - L_1 = \frac{\pi}{4}$

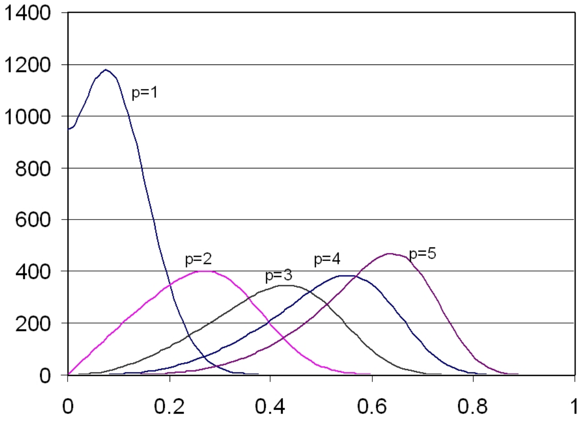

Fig. 9. Value of $|\Theta_{h^2}|$ in the high-order schemes for $k = 8,\ m = 1,\ \sigma = 150,\ L_2 - L_1 = \frac{\pi}{4}$

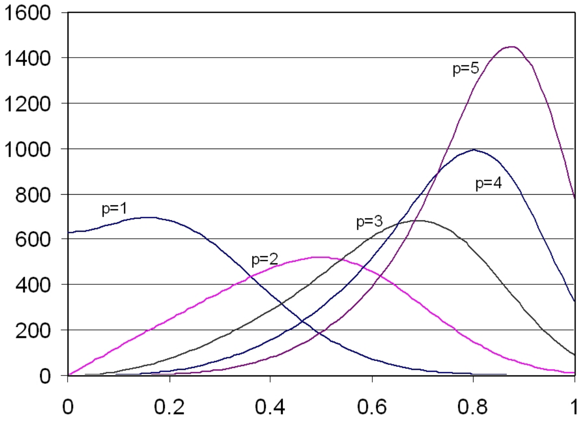

Fig. 10. Value of $|\Theta_{h^2}|$ in the high-order schemes for $k = 8,\ m = 1,\ \sigma = 50,\ L_2 - L_1 = \frac{\pi}{8}$

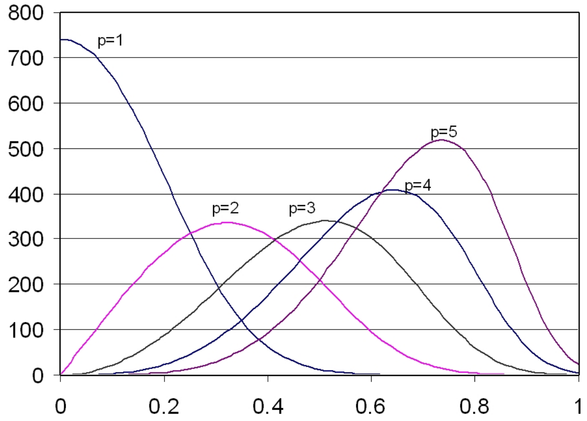

Fig. 11. Value of $|\Theta_{h^2}|$ in the high-order schemes for $k = 12,\ m = 1,\ \sigma = 50,\ L_2 - L_1 = \frac{\pi}{4}$

In Figures 8-11 we find similar results to the ones we found in the case of the second order schemes. We see that the dependence on the value of $(k^2 - m^2)$ is reduced, which means better performance for large values of k (compare Figures 6 and 11). Another quality, which we see from these figures and also from (30) is that $p = 1$ is a very poor choice, because $|\Theta_{h^2}|$ achieves its maximum when $z \to 0^+$. This does not occur when $p > 1$.

7 The modified equation

Another approach to analyze the numerical solution is the modified equation. We start with (25) and (26), find the analytical solution for the equation

$$\frac{\partial}{\partial x}\left(\frac{u_x}{S_x}\right) + \frac{\partial}{\partial y}\left(S_x u_y\right) + k^2 S_x u +$$
$$h^2 \left(\begin{array}{l} \frac{1}{24}\left(\frac{\partial^3}{\partial x^3}\left(\frac{u_x}{S_x}\right) + \frac{\partial}{\partial x}\left(\frac{u_{xxx}}{S_x}\right) + 2 S_x u_{yyyy}\right) \\ +\frac{1-\alpha}{2}\left(\frac{\partial^2}{\partial y^2}\frac{\partial}{\partial x}\left(\frac{u_x}{S_x}\right) + \frac{\partial^2}{\partial x^2}\left(S_x u_{yy}\right)\right) \\ +k^2\left(\beta_s + 2\beta_c\right)\left(\left(S_x u\right)_{xx} + S_x u_{yy}\right) \end{array} \right) = 0$$

When the input boundary condition is $\sin(my)$ then the only Fourier coefficient that does not vanish is $U = \hat{u}\left(x, m\right)$. We recover the ODE

$$\left(\frac{U'}{S_x}\right)' + \left(k^2 - m^2\right) S_x U + \tag{31}$$
$$h^2 \begin{pmatrix} \frac{1}{24}\left(\left(\frac{U'}{S_x}\right)''' + \left(\frac{U'''}{S_x}\right)' + 2m^4 S_x U\right) - \\ \frac{1-\alpha}{2} m^2 \left(\left(\frac{U'}{S_x}\right)' + \left(S_x U\right)''\right) + \\ k^2 \left(\beta_s + 2\beta_c\right)\left(\left(S_x U\right)'' - m^2 S_x U\right) \end{pmatrix} = 0,$$

with the boundary conditions

$$U(0) = 1,\ U(L_2) = 0.$$

Because our main interest is in the high-order schemes we choose

$$\alpha = \frac{5}{6},\ \beta_s + 2\beta_c = \frac{1}{12}$$

(for the $O(h^6)$ we take $\beta_s + 2\beta_c = \frac{1}{12} + O\left(k^2 h^2\right)$). So, (31) becomes

$$\left(\frac{U'}{S_x}\right)' + \left(k^2 - m^2\right) S_x U \tag{32}$$
$$= -\frac{h^2}{24} \begin{pmatrix} \left(\frac{U'}{S_x}\right)''' + \left(\frac{U'''}{S_x}\right)' - \\ 2m^2 \left(\frac{U'}{S_x}\right)' + 2\left(k^2 - m^2\right)\left(\left(S_x U\right)'' - m^2 S_x U\right) \end{pmatrix}.$$

If $p = 1$ then S_x is not differentiable near $x = L_1$. Hence, we assume that $p \geq 2$. In the interior $S_x = 1$ and the right hand side of the equation becomes

$$-\frac{h^2}{12}\left(U^{(4)} - m^2 U'' + \left(k^2 - m^2\right)\left(U'' - m^2 U\right)\right)$$
$$= -\frac{h^2}{12}\left(\frac{d^2}{dx^2}\left(U'' + \left(k^2 - m^2\right) U\right) - m^2\left(U'' + \left(k^2 - m^2\right) U\right)\right)$$

and we can rewrite (32)

$$\left(1 + \frac{h^2}{12}\left(\frac{d^2}{dx^2} - m^2\right)\right)\left(U'' + \left(k^2 - m^2\right) U\right) = 0$$

which leads to

$$U'' + \left(k^2 - m^2\right) U = 0.$$

which is the Fourier expansion of the Helmholtz equation and so in the interior, we solve the Helmholtz equation through $O(h^2)$. Unlike (6) we have not found the exact solution to (32). When $h^2 << 1$ we look for a perturbed solution in the form of

$$U = e^{\pm i\sqrt{k^2 - m^2}\int_0^x S_x(r)\left(1 + h^2 q_{\pm}(r)\right) dr}. \tag{33}$$

Inserting this into (32) and collecting all the terms with $O\left(h^2\right)$ yields a differential equation for $q_\pm$

$$\begin{aligned} & \pm i\sqrt{k^2-m^2}q'_\pm - 2\left(k^2-m^2\right)q_\pm S_x \\ = & \frac{k^2-m^2}{24}\left(\pm 2i\sqrt{k^2-m^2}S'_xS_x + 3S''_x + \frac{1}{\pm i\sqrt{k^2-m^2}}\left(\frac{S''_x}{S_x}\right)'\right). \end{aligned}$$

In the interior the right hand side of this equation vanishes. This emphasizes that only the $O\left(h^2\right)$ factor comes from the PML approximation. Labeling $g_\pm(x) = \pm i\sqrt{k^2-m^2}\int^x S_x(r)\,dr$ and multiplying the equation with the integrating factor $e^{2g_\pm(x)}$ we get

$$\left(q_\pm e^{2g_\pm}\right)' = -\frac{1}{24}\left(2g''_\pm g'_\pm + 3g'''_\pm + \left(\frac{g'''_\pm}{g'_\pm}\right)'\right)e^{2g_\pm}$$

Integrating by parts yields

$$q_\pm = \frac{-1}{24}\left(\pm i\sqrt{k^2-m^2}S'_x + \frac{S''_x}{S_x}\right) + d_\pm e^{\mp 2i\sqrt{k^2-m^2}\int_0^x S_x(r)dr}$$

with constants $d_\pm$. Inserting into (33) we obtain an estimate for the numerical solution using the high-order schemes

$$\hat{u}(x,m) = c_+e^{i\sqrt{k^2-m^2}\int_0^x S_x(r)dr - \frac{h^2}{48}\Psi_+(x)} + c_-e^{-i\sqrt{k^2-m^2}\int_0^x S_x(r)dr - \frac{h^2}{48}\Psi_-(x)}. \quad (34)$$

where $\delta_\pm$ are constants and

$$\begin{aligned} \Psi_+(x) &= -\left(k^2-m^2\right)S_x^2(r)\,|_0^x + i\sqrt{k^2-m^2}2S'_x(r)\,|_0^x + \delta_+e^{-2i\sqrt{k^2-m^2}\int_0^x S_x(r)dr} \\ \Psi_-(x) &= -\left(k^2-m^2\right)S_x^2(r)\,|_0^x - i\sqrt{k^2-m^2}2S'_x(r)\,|_0^x + \delta_-e^{+2i\sqrt{k^2-m^2}\int_0^x S_x(r)dr} \end{aligned}$$

In the interior $S_x = 1$ and so $(S_x^2)'$ and S''_x vanish. So for $x \in [0, L_1]$

$$\hat{u}(x,m) = c_+e^{i\sqrt{k^2-m^2}x - \frac{h^2}{48}\delta_+e^{-2i\sqrt{k^2-m^2}x}} + c_-e^{-i\sqrt{k^2-m^2}x - \frac{h^2}{48}\delta_-e^{2i\sqrt{k^2-m^2}x}}. \quad (35)$$

To find the constants $c_\pm$ and $\delta_\pm$ we use the boundary conditions

$$\hat{u}(0,m) = 1 + 0\cdot h^2 \qquad \hat{u}(L_2,m) = 0 + 0\cdot h^2$$

and the approximation

$$e^{a(x)+h^2b(x)} \simeq e^{a(x)}\left(1 + h^2b(x)\right). \quad (36)$$

Using (34, 35) near the boundaries

$$\hat{u}\left(0,m\right)=c_{+}e^{-\frac{h^2}{48}\delta_{+}}+c_{-}e^{-\frac{h^2}{48}\delta_{-}}\simeq c_{+}\left(1-\frac{h^2}{48}\delta_{+}\right)+c_{-}\left(1-\frac{h^2}{48}\delta_{-}\right)$$

and

$$\begin{aligned}&\hat{u}\left(L_2,m\right)\\&=c_{+}e^{i\sqrt{k^2-m^2}\int_0^{L_2}S_x(r)dr-\frac{h^2}{48}\Psi_{+}(L_2)}+c_{-}e^{-i\sqrt{k^2-m^2}\int_0^{L_2}S_x(r)dr-\frac{h^2}{48}\Psi_{-}(L_2)}\\&\simeq c_{+}e^{i\sqrt{k^2-m^2}\int_0^{L_2}S_x(r)dr}\left(1-\frac{h^2}{48}\Psi_{+}\left(L_2\right)\right)+\\&\quad c_{-}e^{-i\sqrt{k^2-m^2}\int_0^{L_2}S_x(r)dr}\left(1-\frac{h^2}{48}\Psi_{-}\left(L_2\right)\right).\end{aligned}$$

Solving these equations we find that the values of c_{+} and c_{-} are the same as in the formula for u_{F-pml} (11), and

$$\delta_{+}=4i\sqrt{k^2-m^2}S_x'|_0^{L_2}\frac{\eta}{\eta-1}\qquad\delta_{-}=4i\sqrt{k^2-m^2}S_x'|_0^{L_2}\frac{1}{\eta-1}$$

where the value of η is given in (16). Inserting into (35), approximating using (36) and rearranging, we get the approximate solution of the high-order schemes in the interior, which we label u_{N-pml}

$$\begin{aligned}&u_{N-pml}\\&=\left(\begin{array}{c}\frac{-1}{\eta-1}e^{i\sqrt{k^2-m^2}x}+\frac{\eta}{\eta-1}e^{-i\sqrt{k^2-m^2}x}\\-i\frac{h^2}{12}\sqrt{k^2-m^2}S_x'|_0^{L_2}\frac{\eta}{(\eta-1)^2}\left(e^{i\sqrt{k^2-m^2}x}-e^{-i\sqrt{k^2-m^2}x}\right)\end{array}\right)\sin my\\&=u_{F-pml}+\frac{h^2}{6}\sqrt{k^2-m^2}S_x'|_0^{L_2}\frac{\eta}{\left(\eta-1\right)^2}\sin\left(\sqrt{k^2-m^2}x\right)\sin my.\end{aligned}\tag{37}$$

Using our choice for S_x:

$$\begin{aligned}&\left|u_{N-pml}-u_{F-pml}\right|\\&=\frac{h^2}{6}\sqrt{k^2-m^2}\left|\frac{\sigma p}{ik\left(L_2-L_1\right)}\frac{\eta}{\left(\eta-1\right)^2}\sin\left(\sqrt{k^2-m^2}x\right)\right|\\&\simeq\frac{h^2}{6}\frac{\sigma\sqrt{1-\varepsilon}p}{\left(L_2-L_1\right)}e^{-\frac{2\sigma\sqrt{1-\varepsilon}(L_2-L_1)}{p+1}}\end{aligned}\tag{38}$$

One of our goals was to show that the use of the second order approximation inside the PML layer does not corrupt the high-order accuracy in the interior. We can prove this for the $O\left(h^4\right)$ schemes (23) from the error estimate (38).

Taking a large value of $\frac{2\sigma\sqrt{1-\varepsilon}(L_2-L_1)}{p+1}$ then the term $2e^{-\frac{2\sigma\sqrt{1-\varepsilon}(L_2-L_1)}{p+1}}$ is exponentially small and so is negligible compared with $O\left(h^r\right)$ for all integers r. As σ becomes large the exponential term dominates and the accuracy improves. However, we should not choose σ large that σh^2 is large. Thus, in the interior

$$u_{N-pml} = u_{F-pml} + O\left(h^4\right) = u_{exact} + O\left(h^4\right) + O\left(h^2\right) e^{-\frac{2\sigma\sqrt{1-\varepsilon}(L_2-L_1)}{p+1}} \qquad (39)$$

where $O\left(h^4\right)$ comes from the accuracy of the scheme. For the $O\left(h^6\right)$ scheme numerical computations demonstrate a similar error estimate

$$u_{N-pml} = u_{exact} + O\left(h^6\right).$$

8 Numerical results

We present computational results for the Helmholtz equation in the semi-infinite strip. The region to the right of L_1 is replaced by a PML equation which is then truncated at L_2. In each computation we wish to demonstrate one of the properties that has been analyzed. For all the results we use a uniform grid-size, where n denotes the number of gridpoints along the y axis, so $h = \frac{\pi}{n}$. We measure the error in the maximum norm in the square $[0,\pi] \times [0,\pi]$ and so $L_1 \geq \pi$. The numerical approximation is denoted by ϕ. We use a standard LU factorization to solve the linear systems that arise from the finite differences schemes. For fine grids a LU solver is not efficient.

8.1 The case $m > k$: evanescent waves

We wish to verify our error estimate $e^{-\sqrt{m^2-k^2}(2L_2-L_1)}$. For the test case we consider

$$m = 5,\ k = 4.5,\ L_1 = \pi,\ L_2 = \frac{6}{5}\pi,\ p = 2,\ \sigma = 20,\ n = 32. \qquad (40)$$

In this case

$$e^{-\sqrt{m^2-k^2}(2L_2-L_1)} \simeq 6.87 \times 10^{-5}$$

We solve the problem with the $O\left(h^6\right)$ solver (24) with the parameter $\gamma = 1$. The computed solution ϕ approximates u_{N-pml} and satisfies:

$$err = \|u_{exact} - \phi\| = 6.53 \times 10^{-5}$$

Refining the gridsize, taking $n = 64$ we get computationally

$$err = 7.19 \times 10^{-5}$$

which shows that we cannot improve the accuracy of the computation because we reached our desired level of accuracy.

Choosing instead $L_2 = 1.6\pi$ in (40) we get

$$e^{-\sqrt{m^2-k^2}(2L_2-L_1)} \simeq 4.29 \times 10^{-7}$$

and with $n = 32$ we get computationally

$$err = 1.09 \times 10^{-6},$$

with $n = 64$

$$err = 2.01 \times 10^{-7},$$

and with $n = 128$

$$err = 2.13 \times 10^{-7}.$$

Now, the first refinement of the grid improves the result, but we cannot improve it further, as we can see in the second refinement. That is because we again reached our lower bound.

For a second test case we choose default values

$$\begin{aligned} &m=5,\ k=1,\ L_1=\pi,\ L_2=\frac{3}{2}\pi,\ p=2,\ \sigma=20,\\ &n=32, O\left(h^6\right) \text{ scheme with } \gamma=1. \end{aligned}$$

We have the lower error bound

$$e^{-\sqrt{m^2-k^2}(2L_2-L_1)} \simeq 4.28 \times 10^{-14}$$

and choose as our base error

$$err_{base} = 8.18 \times 10^{-7}$$

which is far from the lower error bound. In Table 1 we see the effect of parameter changes from the base case. We list only the changes from the default.

- The independence of the error on the values of p (results 1,2), σ (result 3) and $2L_2 - L_1$(result 10).
  We see (result 9) that if we change m, k in such a way that we preserve the value of $m^2 - k^2$, the error changes. We can explain this by the assumption that as m increases the accuracy of the numerical schemes decreases.
- The efficiency of our high-order scheme for these problems. We can clearly verify the $O\left(h^6\right)$ behavior versus the $O\left(h^4\right)$ one (results 4-8).
- The value of γ in the high order schemes is of minor importance and affects the parameter c in the leading order of the error ch^4 or ch^6.
- We maintain the same behavior even for $k \longrightarrow 0$ (result 11).

no.	parameter changed	value of err
1	$p = 1$	err_{base}
2	$p = 4$	err_{base}
3	$\sigma = 100$	err_{base}
4	$n = 64$	$1.28 \times 10^{-8} \simeq \left(\frac{1}{2}\right)^6 \cdot err_{base}$
5	$n = 128$	$1.99 \times 10^{-10} \simeq \left(\left(\frac{1}{2}\right)^6\right)^2 \cdot err_{base}$
6	$O\left(h^6\right)$ scheme with $\gamma = 3$	8.06×10^{-7}
7	$O\left(h^4\right)$ scheme with $\gamma = 1,\ n = 32$	6.16×10^{-6}
8	$O\left(h^4\right)$ scheme with $\gamma = 1,\ n = 64$	$3.46 \times 10^{-7} \simeq \left(\frac{1}{2}\right)^4 \cdot \#7$
9	$m = 7,\ k = 5$	3.58×10^{-6}
10	$L_1 = \frac{5}{4}\pi,\ L_2 = \frac{13}{8}\pi$	err_{base}
11	$k = 0.0000001$	8.51×10^{-7}

Table 1
The Case $m > k$

In a real problem we cannot control the value of m (the input boundary condition). In the Fourier series expansion we have a complete set of non-zero values. We concentrate on the leading mode value.

8.2 The case $m < k$: traveling waves

8.2.1 The dependence on the accuracy of the scheme

We wish to computationally verify our proof that the global behavior in the interior depends only on the local accuracy used in the interior. We examine the high-order schemes as well as the standard $O\left(h^2\right)$ scheme on a test case. To check this behavior we have to exclude errors that arise from other approximations. Thus, we take in (17) $2e^{-\frac{2\sigma\sqrt{1-\varepsilon}(L_2-L_1)}{p+1}} \longrightarrow 0$ and make sure that $\frac{h^2}{6}\frac{\sigma\sqrt{1-\varepsilon}p}{(L_2-L_1)}e^{-\frac{2\sigma\sqrt{1-\varepsilon}(L_2-L_1)}{p+1}}$ in (38) is negligible.

We take as our base test case

$$m = 6,\ k = 8,\ L_1 = \frac{5}{4}\pi,\ L_2 = \frac{7}{4}\pi,\ p = 2,\ \sigma = 25$$

In this case

$$2e^{-\frac{2\sigma\sqrt{1-\varepsilon}(L_2-L_1)}{p+1}} \simeq 6.03 \times 10^{-8}$$

n	PT	HO-4 $\gamma=0$	HO-4 $\gamma=2$	HO-6 $\gamma=0$	HO-6 $\gamma=2$
16	1.72×10^{0}	5.91×10^{-2}	3.15×10^{-1}	4.25×10^{-2}	1.46×10^{-1}
32	4.85×10^{-1}	4.21×10^{-3}	1.97×10^{-2}	6.86×10^{-4}	2.32×10^{-3}
48	2.17×10^{-1}	8.54×10^{-4}	3.93×10^{-3}	6.98×10^{-5}	2.07×10^{-4}
64	1.23×10^{-1}	2.73×10^{-4}	1.25×10^{-3}	1.49×10^{-5}	3.72×10^{-5}
80	7.86×10^{-2}	1.13×10^{-4}	5.12×10^{-4}	4.74×10^{-6}	9.89×10^{-6}
96	5.46×10^{-2}	5.45×10^{-5}	2.47×10^{-4}	1.93×10^{-6}	3.38×10^{-6}
112	4.02×10^{-2}	2.95×10^{-5}	1.34×10^{-4}	9.07×10^{-7}	1.38×10^{-6}
128	3.08×10^{-2}	1.73×10^{-5}	7.85×10^{-5}	4.71×10^{-7}	6.47×10^{-7}

Table 2
$err = \|u_{exact} - \phi\|_{\infty}$ for $m = 6,\ k = 8,\ L_1 = \frac{5}{4}\pi,\ L_2 = \frac{7}{4}\pi,\ p = 2,\ \sigma = 25$

and

$$\frac{\sigma\sqrt{1-\varepsilon p}}{6\left(L_2 - L_1\right)} \simeq 3.5$$

and if $n \geq 16$

$$h^2 = \left(\frac{\pi}{n}\right)^2 \leq \left(\frac{\pi}{16}\right)^2 \simeq 3.8\times10^{-2}$$

and

$$h^2\frac{\sigma\sqrt{1-\varepsilon p}}{6\left(L_2 - L_1\right)} \leq 0.14\ .$$

Thus, our high-order schemes should yield good results. In Table 2 we see the value of $err = \|u_{exact} - \phi\|_{\infty}$ resulting from the use of the schemes: PT (22), HO-4 (23) and HO-6 (24). Increasing the value of n decreases the value of the gridsize h in both directions. We assume that for small values of h,

$$err \simeq C\left(k\right)n^{-r} = \frac{C\left(k\right)}{\pi^r}h^r$$

where r is the order of the scheme, i.e.

$$r = \begin{cases} 2 & \text{for PT scheme} \\ 4 & \text{for the schemes EB,HO} \\ 6 & \text{for the schemes EB-6,HO-6} \end{cases}$$

Hence, if $n = 2^l$

$$-\log_2\left(err\right) \simeq l\cdot r - \log_2 C\left(k\right).$$

We calculate r, the order of accuracy, by measuring the slope of the curve (in figure 12). This computationally verifies our hypothesis.

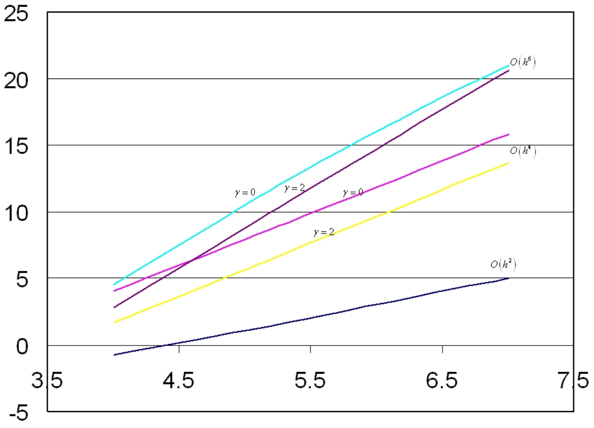

Fig. 12. $-\log(err)$ by $\log(n)$ for the finite differences schemes

The computing time required for all of these scheme, applied on a given problem, is the same when we use Gaussian elimination. This fact emphasizes the benefit of using the high order schemes.

8.2.2 Convergence to the modified solution

We wish to confirm the modified equation's solution (37) which we labeled u_{N-pml}. We start with the $O\left(h^4\right)$ approximation with $\gamma = 0$, and solve the problem with

$$m = 1,\ k = 5,\ L_1 = \frac{5}{4}\pi,\ L_2 = \frac{3}{2}\pi,\ p = 2,\ \sigma = 20.$$

In this case the lower error bound is rather poor and we get

$$2e^{-\frac{2\sigma\sqrt{1-\varepsilon}(L_2-L_1)}{p+1}} \simeq 7 \times 10^{-5}.$$

Let ϕ be the computational approximation, u_{exact} be the exact solution to the combined Helmholtz-PML problem and u_{NPML} given by (37). We measure two kinds of errors

$$err_{ex} = \left\|u_{exact} - \phi\right\|_{\infty}$$

and

$$err_{ho} = \left\|u_{N-pml} - \phi\right\|_{\infty}$$

Our goal is to show that the scheme converges to the u_{N-pml} solution with $O\left(h^4\right)$. We collect the results in Table 3. We see that the scheme approximates the u_{N-pml} solution and

$$\frac{err_{ho}\left(n=32\right)}{err_{ho}\left(n=16\right)} \simeq \frac{err_{ho}\left(n=64\right)}{err_{ho}\left(n=32\right)} \simeq \frac{err_{ho}\left(n=128\right)}{err_{ho}\left(n=64\right)} \simeq \left(\frac{1}{2}\right)^{-4}$$

n	err_{ex}	err_{ho}
16	3.05×10^{-2}	3.04×10^{-2}
32	2.00×10^{-3}	1.93×10^{-3}
64	1.85×10^{-4}	1.19×10^{-4}
128	7.67×10^{-5}	7.41×10^{-6}

Table 3
Convergence to u_{ho} for $m=1, k=5, L_1=\frac{5}{4}\pi, L_2=\frac{3}{2}\pi, p=2, \sigma=20$ with HO-4 scheme

n	err_{ex}	err_{ho}
16	6.19×10^{-3}	6.12×10^{-3}
32	5.64×10^{-4}	4.93×10^{-4}
64	9.91×10^{-5}	2.88×10^{-5}
128	7.18×10^{-5}	1.76×10^{-6}

Table 4
Convergence to u_{ho} for $m=1, k=5, L_1=\frac{5}{4}\pi, L_2=\frac{3}{2}\pi, p=2, \sigma=20$ with HO-6 scheme

n	$err_{ex} = err_{ho}$
16	1.81×10^{-3}
32	6.29×10^{-5}
64	3.33×10^{-6}
128	1.99×10^{-7}

Table 5
Convergence to u_{ho} for $m=1, k=5, L_1=\frac{5}{4}\pi, L_2=2\pi, p=2, \sigma=20$ with HO-6 scheme

which is consistent with our analysis. The analytic solution is u_{exact}. To converge to u_{exact} we set the parameters such that the exponent $2e^{-\frac{2\sigma\sqrt{1-\varepsilon}(L_2-L_1)}{p+1}}$ is small enough.

We solve the same problem with the $O\left(h^6\right)$ scheme with $\gamma = 0$ (results in Table 4). Again, we see that we approach u_{N-pml} and not u_{exact}, but this time we do not get a $O\left(h^6\right)$ approximation. For instance $\frac{err_{ho}(n=64)}{err_{ho}(n=32)} \simeq \left(\frac{1}{2}\right)^{-4}$ and not $\left(\frac{1}{2}\right)^{-6}$. To find the cause we change L_2 to 2π, to achieve better damping in the PML. (Table 5). This time $\frac{err_{ho}(n=64)}{err_{ho}(n=32)} \simeq \left(\frac{1}{2}\right)^{-4.48}$, and in other problems (last subsection) we do get $O\left(h^6\right)$. The problem is that the evaluated u_{N-pml} is not what we are approximating in the $O\left(h^6\right)$ scheme. In the analysis of the modified equation we approximated the $O\left(h^2\right)$ term and neglected the $O\left(h^4\right)$

n	$err_{ex} = err_{ho}$
16	4.77×10^{-3}
32	1.15×10^{-5}
64	1.87×10^{-7}
128	2.98×10^{-9}

Table 6
Convergence to u_{ho} for $m = 1, k = 5, L_1 = \frac{5}{4}\pi, L_2 = \frac{7}{4}\pi, p = 4, \sigma = 200$ with HO-6 scheme

term. However, this high order term is significant in some problems and in others is negligible. Choosing $p = 2$ can be accurate for some problems (see Table 2), however, we should choose $p \geq 4$ for the $O\left(h^6\right)$ scheme. For instance in Table 6 , we do get the desired approximation

$$\frac{err_{ho}\,(n=64)}{err_{ho}\,(n=32)} \simeq \frac{err_{ho}\,(n=128)}{err_{ho}\,(n=64)} \simeq \left(\frac{1}{2}\right)^{-6}.$$

8.2.3 Selecting the parameters in the function σ_x

We wish to check the influence of the parameters we choose in the approximation, $L_1,\ L_2 - L_1,\ p$ and σ, on a given problem.

8.2.3.1 Value of L_1 : The error estimates we obtained imply that the value of L_1 is relatively unimportant (we approximated $|\frac{2}{\eta-1}|$ that depends on L_1 by $|\frac{2}{\eta}|$). We check this with the test case

$$\begin{gathered} m=1,\ k=10,\ L_1 - L_2 = \frac{\pi}{4},\ \ p=4,\ \sigma=100, \\ n=32,\ O\left(h^6\right) \text{ scheme with } \gamma=0. \end{gathered}$$

The value of $\left|\frac{2}{\eta}\right| = 2e^{-\frac{2\sigma\sqrt{1-\varepsilon}(L_2-L_1)}{p+1}} \simeq 5.36 \times 10^{-14}$. We summarize the results in Table 7.

8.2.3.2 Width of the PML, $L_2 - L_1$: To support our theory we show that the accuracy increases as we take a thicker PML layer. Taking

$$m=1,\ k=10,\ L_1 = \frac{5\pi}{4},\ \ p=4,\ \sigma=100,\ O\left(h^6\right) \text{ scheme with } \gamma=0, \quad (41)$$

L_1	$\left\lvert\frac{2}{\eta-1}\right\rvert$	err
1	4.91×10^{-14}	3.01×10^{-3}
1.25	4.73×10^{-14}	3.33×10^{-3}
1.5	4.53×10^{-14}	2.85×10^{-3}
1.75	4.30×10^{-14}	3.44×10^{-3}
2	4.03×10^{-14}	2.68×10^{-3}
2.25	3.75×10^{-14}	3.54×10^{-3}
2.5	3.44×10^{-14}	2.50×10^{-3}

Table 7
The error dependance on the value of L_1

$L_2 - L_1$	$2e^{-\frac{2\sigma\sqrt{1-\varepsilon}(L_2-L_1)}{p+1}}$	err $n = 32$	err $n = 64$
$\frac{\pi}{16}$	8.08×10^{-4}	3.13×10^{-1}	8.53×10^{-3}
$\frac{\pi}{8}$	3.26×10^{-7}	5.92×10^{-2}	6.08×10^{-5}
$\frac{\pi}{4}$	5.31×10^{-14}	3.33×10^{-3}	3.34×10^{-5}
$\frac{\pi}{2}$	1.41×10^{-27}	2.16×10^{-3}	3.33×10^{-5}
π	9.98×10^{-55}	2.16×10^{-3}	3.33×10^{-5}

Table 8
The error in the HO-6 scheme as a function of the width of the PML

$L_2 - L_1$	$k = 2.5$	$k = 5$	$k = 13$
$\frac{\pi}{16}$	1.68×10^{-1}	2.42×10^{-1}	4.21×10^{-1}
$\frac{\pi}{8}$	1.34×10^{-2}	2.16×10^{-2}	1.10×10^{-1}
$\frac{\pi}{4}$	5.16×10^{-5}	1.78×10^{-4}	1.26×10^{-2}
$\frac{\pi}{2}$	1.03×10^{-7}	1.23×10^{-5}	1.44×10^{-2}
π	1.80×10^{-8}	1.24×10^{-5}	1.44×10^{-2}

Table 9
The error in the HO-6 scheme as a function of width of PML and k, for $n = 32$

we summarize the results in Table 8. We verify that as the mesh becomes finer we maintain the $O\left(h^6\right)$ accuracy if the PML is thick enough. Increasing the width of the PML beyond a certain limit does not improve the accuracy while requiring more storage and computational time. For small values of h we obtain more accurate results, but need a thicker PML as seen in Table 9.

Our main interest is to take a thin PML, while maintaining the accuracy. We control this by increasing the value of σ. In Table 10 we set $2e^{-\frac{2\sigma\sqrt{1-\varepsilon}(L_2-L_1)}{p+1}} =$

$L_2 - L_1$	σ	err	err
		$n = 32$	$n = 64$
$\frac{\pi}{8}$	400	1.7×10^{-1}	9.73×10^{-4}
$\frac{\pi}{4}$	200	5.04×10^{-3}	3.36×10^{-5}
$\frac{\pi}{2}$	100	2.16×10^{-3}	3.33×10^{-5}
π	50	2.16×10^{-3}	3.33×10^{-5}

Table 10
The error in the HO-6 scheme for a constant value of $2e^{-\frac{2\sigma\sqrt{1-\varepsilon}(L_2-L_1)}{p+1}}$

n	σ	$L_2 - L_1$	HO-4	HO-6
16	18.75	$\frac{\pi}{2}$	9.66×10^{-1}	1.40×10^{-1}
32	37.5	$\frac{\pi}{4}$	5.15×10^{-2}	3.02×10^{-3}
64	75	$\frac{\pi}{8}$	3.62×10^{-4}	9.16×10^{-4}

Table 11
Results for 8 points inside the PML with $m\!=\!2, k\!=\!10, L_1\!=\!\frac{5\pi}{4}, p\!=\!2, \frac{\sigma(L_2-L_1)}{p+1} = \frac{25\pi}{8}$

n	σ	$L_2 - L_1$	HO-4	HO-6
16	31.25	$\frac{\pi}{2}$	1.06×10^{0}	1.08×10^{-1}
32	62.5	$\frac{\pi}{4}$	5.14×10^{-2}	1.79×10^{-3}
64	125	$\frac{\pi}{8}$	3.14×10^{-3}	7.83×10^{-5}

Table 12
Results for 8 points inside the PML $m\!=\!2, k\!=\!10, L_1\!=\!\frac{5\pi}{4}, p\!=\!4, \frac{\sigma(L_2-L_1)}{p+1} = \frac{25\pi}{8}$

1.41×10^{-27} for the test problem (41). We conclude that as long as the PML is not too thin and we have enough points in the PML, we get the desired accuracy.

8.2.3.3 Fixed number of points in the PML: Computational practice, is to set a fixed number of points inside the PML. Thus, when we decrease the gridsize, we change the physical width of the PML. In the above computations we choose the physical width of the PML as constant. We want to see if this influences our results. To check the behavior we set a problem with

$$m = 2,\ k = 10,\ L_1 = \frac{5\pi}{4}$$

and for different values of gridsize h, use the same number of points inside the PML. In the first test (tables 11 - 13) we use 8 points in the PML, and in the second test (Table 14 and 15) 16 points. We see that we lose the $O\left(h^6\right)$ accuracy when $p = 2$ (tables 11 and 14), but maintain the $O\left(h^4\right)$ behavior.

n	σ	$L_2 - L_1$	HO-4	HO-6
16	18.75	$\frac{\pi}{2}$	1.01×10^{0}	1.34×10^{-1}
32	37.5	$\frac{\pi}{4}$	5.16×10^{-2}	1.64×10^{-3}
64	75	$\frac{\pi}{8}$	3.17×10^{-3}	3.06×10^{-5} (2.55×10^{-5} towards u_{N-pml})

Table 13
Results for 8 points inside the PML. $m=2, k=10, L_1=\frac{5\pi}{4}, p=4, \frac{\sigma(L_2-L_1)}{p+1}=\frac{15\pi}{8}$

n	σ	$L_2 - L_1$	HO-4	HO-6
16	9.375	π	9.94×10^{-1}	1.23×10^{-1}
32	18.75	$\frac{\pi}{2}$	5.16×10^{-2}	1.67×10^{-3}
64	37.5	$\frac{\pi}{4}$	3.21×10^{-3}	1.19×10^{-4}

Table 14
Results for 16 points inside PML with $m=2, k=10,\ L_1=\frac{5\pi}{4}, p=2, \frac{\sigma(L_2-L_1)}{p+1}=\frac{25\pi}{8}$

n	σ	$L_2 - L_1$	HO-4	HO-6
16	15.625	π	9.94×10^{-1}	1.24×10^{-1}
32	31.25	$\frac{\pi}{2}$	5.17×10^{-2}	1.65×10^{-3}
64	62.5	$\frac{\pi}{4}$	3.17×10^{-3}	2.52×10^{-5}

Table 15
Results for 16 points inside the PML with $m=2, k=10, L_1=\frac{5\pi}{4}, p=4, \frac{\sigma(L_2-L_1)}{p+1}=\frac{25\pi}{8}$

This is because for the $O\left(h^6\right)$ scheme we have to choose $p \geq 4$ and have a sufficiently wide layer (compare with the results in the last section). For thin PML layers it is noticed (Table 12) that we cannot maintain the $O\left(h^6\right)$ behavior if we choose a large σ, but by taking smaller value of σ we lose the approximation to u_{exact} and maintain the $O\left(h^6\right)$ accuracy towards the modified equation solution u_{N-pml}. Thus, if we wish to work with a fixed number of points inside the PML we should set the parameters such that $2e^{-\frac{2\sigma\sqrt{1-\varepsilon}(L_2-L_1)}{p+1}}$ is small enough and choose $p \geq 4$ for HO-6 and $p = 2$ for HO-4.

We denote *n_pml* as the number of points in the PML

$$2e^{-\frac{2\sigma\sqrt{1-\varepsilon}(L_2-L_1)}{p+1}} = 2e^{-\frac{2h\cdot n\_pml\cdot\sigma\sqrt{1-\varepsilon}}{p+1}}.$$

If we know the minimal value of h *á* priori (In tables 11-15 it is $\frac{\pi}{64}$), we can set *n_pml* and σ to satisfy a desired estimate for $2e^{-\frac{2h\cdot n\_pml\cdot\sigma\sqrt{1-\varepsilon}}{p+1}}$. Hence, when we refine the interior grid to increase the accuracy we also need more points in the PML.

k	8	7	6.5	6.25	6.1	6.05
error	6.86 10^{-04}	2.44 10^{-04}	1.16 10^{-04}	.00138	.0182	.069

Table 16
$m = 6$ and k approaches m with $L_1 = \frac{5\pi}{4}, L_2 = \frac{7\pi}{4}, \sigma = 25$ with 32 points

p	estimate (17)	err $n = 32$	err $n = 64$	err $n = 128$	$\frac{err_{32}}{err_{64}}$	$\frac{err_{64}}{err_{128}}$
1	1.32×10^{-68}	1.19×10^{-1}	2.08×10^{-2}	4.86×10^{-3}	5.7	2.1
2	1.12×10^{-45}	1.98×10^{-3}	7.70×10^{-5}	7.59×10^{-6}	25.7	25.1
3	2.30×10^{-34}	2.22×10^{-3}	3.50×10^{-5}	6.21×10^{-7}	63.4	56.3
4	1.41×10^{-27}	2.16×10^{-3}	3.33×10^{-5}	5.21×10^{-5}	64.9	63.9
5	4.74×10^{-23}	2.16×10^{-3}	3.33×10^{-5}	5.21×10^{-5}	64.9	63.9
10	4.56×10^{-13}	2.02×10^{-3}	3.33×10^{-5}	5.21×10^{-5}	61.2	63.9
30	8.35×10^{-5}	7.23×10^{-2}	5.24×10^{-4}	7.69×10^{-4}	138.0	< 1

Table 17
Results for different values of p for $m = 1, k = 10, L_1 = \frac{5\pi}{4}, L_2 = \frac{7\pi}{4}, \sigma = 100, O\left(h^6\right)$ with $\gamma = 0$

As m and k get closer we approach the evanescent limit and the PML is less effective. The error is shown in table (16).

8.2.3.4 Value of p : We solve the test problem with various values of p (Table 17) with

$$m = 1,\ k = 10,\ L_1 = \frac{5\pi}{4},\ \ L_2 = \frac{7\pi}{4},\ \sigma = 100,\ O\left(h^6\right) \text{ scheme with } \gamma = 0.$$

We clearly see the bad behavior of the case $p = 1$ and the need for sufficient derivatives in S_x as seen in Figures 8-11. In the $O\left(h^6\right)$ case we should choose $p \geq 4$. Moreover, we do not benefit from taking larger values of p than four.

8.2.3.5 Value of σ : Taking as a test problem with various values of σ,

$$m = 1,\ k = 10,\ L_1 = \frac{5\pi}{4},\ \ L_2 = \frac{3\pi}{2},\ p = 4,\ O\left(h^6\right) \text{ scheme with } \gamma = 0.$$

We present the results in Table 18. We conclude that the value of σ does not

σ	estimate (17)	err $n=32$	err $n=64$	err $n=128$	$\frac{err_{32}}{err_{64}}$	$\frac{err_{64}}{err_{128}}$
10	8.77×10^{-2}	9.54×10^{-2}	9.23×10^{-2}	9.16×10^{-2}	1.0	1.0
20	3.85×10^{-3}	4.59×10^{-3}	3.94×10^{-3}	3.88×10^{-3}	1.2	1.0
40	7.43×10^{-6}	2.03×10^{-3}	3.39×10^{-5}	7.43×10^{-6}	59.8	4.6
80	2.75×10^{-11}	3.24×10^{-3}	3.34×10^{-5}	5.21×10^{-7}	97	64.1
160	3.81×10^{-22}	9.28×10^{-3}	3.35×10^{-5}	5.21×10^{-7}	277	64.3
320	7.25×10^{-44}	1.20×10^{-2}	3.42×10^{-5}	5.30×10^{-7}	350	64.5
10^3	1.31×10^{-136}	3.63×10^{-2}	4.21×10^{-5}	5.89×10^{-7}	86.6	71.5

Table 18
Results for different values of σ for $m{=}1, k{=}10,\ L_1{=}\frac{5\pi}{4}, L_2{=}\frac{3\pi}{2}, p{=}4,\ O\left(h^6\right)\ \gamma{=}0$

g	-10	-6	-4	-3	-1	0
err	2.73 10^{-3}	6.75 10^{-4}	6.66 10^{-4}	6.70 10^{-4}	6.81 10^{-4}	6.86 10^{-4}
g	1	3	4	5	6	10
err	6.87 10^{-4}	6.83 10^{-4}	6.76 10^{-4}	6.68 10^{-4}	6.72 10^{-4}	2.72 10^{-3}

Table 19
Error varying g in $S_x{=}1+\frac{\sigma_x}{g+ik}$. $L_1{=}\frac{5\pi}{4}, L_2{=}\frac{7\pi}{4},\ m{=}6, k{=}8$, 32 points, $\sigma{=}\ 25$

influence the accuracy of the approximation as long as $2e^{-\frac{2\sigma\sqrt{1-\varepsilon}(L_2-L_1)}{p+1}}$ is small enough. Thus, a natural choice of σ should be $\sigma > k$ such that

$$2e^{-\frac{2\sigma\sqrt{1-\varepsilon}(L_2-L_1)}{p+1}} \tilde{\ } 10^{-q}$$

where q is the number of the desired significant digits accuracy. We choose the smallest σ that satisfies this.

We next consider a more general formula for S_x given by $S_x = 1+\frac{\sigma_x}{g+ik}$. The original formula corresponds to $g=0$. In table (19) we present the error as a function of g. We consider $L_1=\frac{5\pi}{4}, L_2=\frac{7\pi}{4}$, $m=6, k=8$ with 32 points and $\sigma=25$. We see that indeed the smallest error occurs for $g=-5.0$. However, the differences are so small that there is no practical purpose in trying to choose a nonzero g. Hence, we continue to use the original $g=0$.

9 Iterative Techniques

For very fine meshes in two dimensions and coarser meshes in three dimensions an LU decomposition becomes very expensive. Hence, we consider the use of an iterative solver. We examine Krylov subspaces methods such as CGNR, GMRES, QMR and BiCG on the combined problem and found that they converge to the solution very slowly. Hence, we need to use a preconditioner for the problem. A preconditioner for the pure Helmholtz equation was constructed in [3]. The preconditioner was based on the inverse of the Laplacian operator computed with one sweep of SSOR.

The major difficulty of this preconditioner on the combined problem is that the PML introduces a different equation. $k = 0$ in the PML does not give the Laplace equation. Instead, we construct a preconditioner, that will act as the inverse Laplacian operator in the interior, and behave as an approximate inverse in the PML. We do not calculate M^{-1} but we approximate the solution of $My = z$, by applying a few sweeps of SSOR or damped Jacobi (DJ). In the PML we cannot choose $k = 0$ because of the definition of S_x. For the Helmholtz equation with small but nonzero k the operator is still positive definite. Thus, to apply the preconditioner M we use the standard approximation of the combined problem with a small value of k which we denote $\tilde{k}$.

$$\begin{aligned}
&\alpha D_x \left(A u_x\right)_j + \frac{1-\alpha}{2}\left(D_x \left(A u_x\right)_{j+1} + D_x \left(A u_x\right)_{j-1}\right) + \\
&\alpha D_y \left(B u_y\right)_i + \frac{1-\alpha}{2}\left(D_y \left(A u_y\right)_{i+1} + D_y \left(A u_y\right)_{i-1}\right) + \\
&\left(1 - 4\beta_s - 4\beta_c\right) C_{i,j} u_{i,j} \\
&+\beta_s \begin{pmatrix} C_{i+1,j} u_{i+1,j} + C_{i-1,j} u_{i-1,j} + \\ C_{i,j+1} u_{i,j+1} + C_{i,j-1} u_{i,j-1} \end{pmatrix} \\
&+\beta_s \begin{pmatrix} C_{i+1,j+1} u_{i+1,j+1} + C_{i-1,j+1} u_{i-1,j+1} + \\ C_{i+1,j-1} u_{i+1,j-1} + C_{i-1,j-1} u_{i-1,j-1} \end{pmatrix} .
\end{aligned}$$

Where $A = \frac{1}{S_x}$, $B = S_x$, $C = \tilde{k}^2 S_x$. The resulting matrix $\hat{M}$ is complex and cannot be applied as a preconditioner for two reasons. It is not symmetric and it is not real and positive definite. It does have these two properties in the interior, but not in the PML.

To overcome these two difficulties we set the PT values: $\alpha = 1$, $\beta_s = \beta_c = 0$ in the preconditioner. This choice makes the approximation complex-symmetric. It is also useful because it involves only 5 unknowns in each equation instead of 9. To change the matrix to a real symmetric positive definite matrix we make the change:

$$M_{i,j} = \begin{cases} \left|\hat{M}_{i,j}\right| & i = j \\ -\left|\hat{M}_{i,j}\right| & i \neq j \end{cases} \qquad \text{for every } 1 \leq i, j \leq N.$$

n	$\tilde{\sigma} = 10$ $\tilde{p} = 1$	$\tilde{\sigma} = 10$ $\tilde{p} = 2$	$\tilde{\sigma} = 1$ $\tilde{p} = 4$	$\tilde{\sigma} = 10$ $\tilde{p} = 4$	$\tilde{\sigma} = 100$ $\tilde{p} = 4$	$\tilde{\sigma} = 10$ $\tilde{p} = 6$
16	211	206	208	200	204	204
32	832	811	927	807	807	871
64	4022	3911	4683	3766	3857	4421

Table 20
Number of iterations in the Left Preconditioned CGNR with diagonal scaling using 4 sweeps of DJ with $\omega = 0.7$ applied on the combined preconditioner with $\tilde{k} = 0.1$.

n	with diagonal scaling	without diagonal scaling
16	351	877
32	1732	12823
64	8646	116744

Table 21
Number of iterations non preconditioned CGNR with and without diagonal scaling

In the interior $M_{i,j} = \hat{M}_{i,j}$. Because $\tilde{k}$ is small in the interior, the approximation is a small perturbation of the Laplacian. We call this preconditioner the combined preconditioner. Note, that as in the pure Laplacian operator preconditioner, this preconditioner corresponds to a second order accurate operator even when the scheme in the interior is higher order accurate.

Table 20 shows the number of iterations for the CGNR preconditioned algorithm with the new combined preconditioner. We choose $\tilde{k} = 0.1$. The best value of p in the preconditioner is the same as used for the approximation. $\tilde{\sigma}$ should be small, but not too small. The benefit of the preconditioner is demonstrated comparing the convergence with that of the non-preconditioned algorithm in Table 21. As shown in [3] as k gets larger the preconditioner is less effective. This can be improved as in [15].

10 Conclusions

We construct a PML for the two dimensional Helmholtz equation in a strip. We prove that the combined Helmholtz and PML layer has a unique solution and that this solution is a good approximation to the interior Helmholtz equation. We find a bound for this error and show that one can control the error by changing the PML parameters. We describe high-order finite differences schemes for the combined Helmholtz PML equation and apply them on the strip problem. We analyze the numerical error governed by the use of these schemes and show that for a proper choice of parameters, the combined scheme maintains the high-order accuracy of the interior approximation. A variety of numerical results is presented to support the analysis.

We also investigate the use of iterative methods to solve the resultant linear equations. We develop a preconditioner Krylov algorithms to solve the combined problem. We conclude that the parameters should be chosen so that:

- The length of the interior, L_1, should be as small as the physical problem permits because it does not influence the accuracy but has a significant impact on the size of the linear system.
- The width of the PML, $L_2 - L_1$, should be as thin as possible to decrease the size of the linear system. Numerical tests show that we cannot make the layer too thin relative to the interior grid. For the interior grids used in this study we could use 16 grid points in the PML and still get high order accuracy. If the grid is refined than one needs more points in the PML.
- The degree of the polynomial in the PML, p, (see (14)) should be 2 for the fourth order accurate scheme and 4 for the sixth order accurate scheme.
- σ, the maximum of $\sigma_x(x)$ (see (14)) should be set to the lowest value which maintains the desired accuracy determined by $e^{-\frac{2\sigma\sqrt{1-\varepsilon}(L_2-L_1)}{p+1}}$ (see (17)).
- If both travelling and evanescent waves are present then the size of the PML should be chosen to damp the evanescent waves. Then the other parameters are chosen to efficiently damp the traveling waves. The difficult situation is travelling waves near resonance.

References

[1] S. Abarbanel and D. Gottlieb, *A Mathematical Analysis of the PML Method*, J. Comput. Physics, 134, 357-363 (1997).

[2] A. Bayliss, C.I. Goldstein and E.Turkel, *An Iterative Method for the Helmholtz Equation*, J. Comput. Physics 49, 443-457 (1983).

[3] A. Bayliss, C.I. Goldstein and E.Turkel, *Preconditioned Conjugate-Gradient Methods for the Helmholtz Equation*, Elliptic Problem Solvers II , 233-243 (1984).

[4] A. Bayliss, M. Gunzburger, E. Turkel, *Boundary Conditions for the Numerical Solution of Elliptic Equations in Exterior Regions*, SIAM J. Appl. Math. 42, 430-451 (1982).

[5] J-P Berenger, *A Perfectly Matched Layer for the Absorption of Electromagnetic Waves*, J. Comput. Physics, 114, 185-200 (1994).

[6] J-P Berenger, *Three Dimensions Perfectly Matched Layer for the Absorption of Electromagnetic Waves*, J. Comput. Physics, 127, 363-379 (1996).

[7] A. Greenbaum, *Iterative Methods for Solving Linear Systems*, SIAM, Philadelphia (1997).

[8] S. Gedney, *The Perfectly Matched Layer Absorbing Media*, Advances in Computational Electrodynamics: The Finite-Difference Time-Domain Method, A. Taflove editor, Artech House 263-343 (1998).

[9] I. Harari and E. Turkel, *Accurate Finite Difference Methods for Time-Harmonic Wave Propagation*, J. Comput. Physics, 119, 252-270 (1995).

[10] P.G. Petropoulos, *On the Termination of the Perfectly Matched Layer with Local Absorbing Boundary Conditions*, J. Comput. Physics, 143, 665-673 (1968).

[11] I. Singer and E. Turkel, *High Order Finite Difference Methods for the Helmholtz Equation*, Comput. Methods Appl. Mech. Eng. 163, 343-358 (1998).

[12] I. Singer and E. Turkel, *Sixth Order Accurate Finite Difference Schemes for the Helmholtz Equation* submitted to Journal of Computational Acoustics.

[13] S. Tsynkov and E. Turkel, *A Cartesian Perfectly Matched Layer for the Helmholtz Equation,* Artificial Boundary Conditions with Applications to CEM, Loic Tourrette editor, Novascience Publishing (2001).

[14] E. Turkel, *Numerical Difficulties Solving Time Harmonic Equations*, Multiscale Computational Methods in Chemistry and Physics, A. Brandt, J. Bernholc and K. Binder editors, IOS Press, Ohmsha, 319-337 (2001).

[15] C. Vuik, Y.A. Erlangaa and C.W. Oosterlee. *Shifted Laplace preconditioners for the Helmholtz Equation*, ENUMATH 2003.

[16] E. Turkel and A. Yefet, *Absorbing PML Boundary Layers for Wave-Like Equations,* Appl. Numer. Math. 27, 533-557 (1998).