Journal of Computational and Applied Mathematics 236 (2011) 1240–1258

Contents lists available at SciVerse ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

An optimal 25-point finite difference scheme for the Helmholtz equation with PML☆

Zhongying Chen, Tingting Wu*, Hongqi Yang

Guangdong Province Key Laboratory of Computational Science, Sun Yat-Sen University, Guangzhou 510275, PR China

ARTICLE INFO

Article history:
Received 24 December 2010
Received in revised form 15 August 2011

Keywords:
Helmholtz equation
PML
25-point finite difference scheme
Numerical dispersion

ABSTRACT

In this paper, we present an optimal 25-point finite difference scheme for solving the Helmholtz equation with perfectly matched layer (PML) in two dimensional domain. Based on minimizing the numerical dispersion, we propose the refined choice strategy for choosing optimal parameters of the 25-point finite difference scheme. Numerical experiments are given to illustrate the improvement of the accuracy and the reduction of the numerical dispersion.

1. Introduction

In a Cartesian coordinate system, the 2-D scalar wave equation with no damping in the frequency domain is given by

$$\Delta u + k^2 u = g, \tag{1.1}$$

where u is the Fourier component of the wavefield pressure, $k \in \mathbb{R}$, called wavenumber, is a positive number, and g is the Fourier transformation of the source function. The above equation, which is the well-known Helmholtz equation, appears directly or indirectly in almost all wave-related problems arisen from many science, engineering, and industry applications. Therefore, solving the Helmholtz equation, in various forms, has always been a hot topic in wave computation (cf. [1,2] and the reference therein).

To compute the solution of the above problem, due to finite memory and speed limitation of the computer, we first need to formulate the problem as a finite domain problem. Thanks to absorbing boundary conditions, we are able to truncate the infinite domain into a finite domain with less or almost no reflection; see one-way approximation (cf. [3–5]), PML (perfectly matched layer, cf. [6–12]), and so on. In this paper, PML is used to truncate the domain and absorb the outgoing waves. The PML was proposed by Bérenger in 1994 in [6], and it has the astonishing property of generating almost no reflection in theory at the interface between the interior medium (the interested domain) and the artificial absorbing medium.

Specifically, applying the PML technique to truncate the infinite domain into a bounded rectangular domain and taking the same notations as in [13] lead to the equation

$$\frac{\partial}{\partial x}\left(A\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left(B\frac{\partial u}{\partial y}\right) + Ck^2 u = f, \tag{1.2}$$

where

$$f := \begin{cases} 0, & \text{inside PML,} \\ g, & \text{outside PML,} \end{cases}$$

☆ This research is partially supported by the Natural Science Foundation of China under grants 10771224 and 11071264, the Science and Technology Section of SINOPEC and Guangdong Provincial Government of China through the "Computational Science Innovative Research Team" program.

* Corresponding author.
E-mail addresses: lnsczy@mail.sysu.edu.cn (Z. Chen), wttxrm@126.com (T. Wu), mcsyhq@mail.sysu.edu.cn (H. Yang).

0377-0427/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2011.08.007

and

$$A := \frac{e_y}{e_x}, \qquad B := \frac{e_x}{e_y}, \qquad C := e_x e_y,$$

in which $e_x := 1 - i\frac{\sigma_x}{\omega}$, $e_y := 1 - i\frac{\sigma_y}{\omega}$, ω is the angular frequency, σ_x is a function only of x defined as

$$\sigma_x := \begin{cases} 2\pi a_0 f_0 \left(\dfrac{l_x}{L_{\mathrm{PML}}}\right)^2, & \text{inside PML,} \\ 0, & \text{outside PML,} \end{cases} \tag{1.3}$$

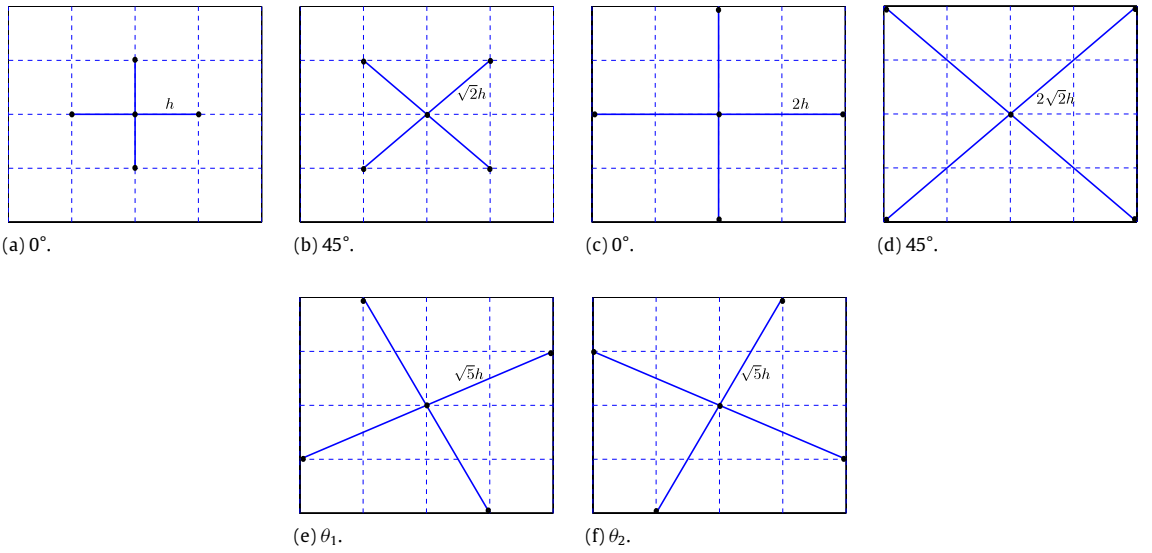
with the dominant frequency f_0 of the source, the thickness L_{PML} of PML, the distance l_x from the interface between the interior region and PML region, and a constant a_0, and σ_y is defined similarly. Eq. (1.2) can be seen as a general form of the Helmholtz equation (1.1) with the corresponding PML equation, since in the interior domain, $A = B = C = 1$. We call it the Helmholtz-PML equation. Moreover, Eq. (1.2) is also valid for variable $k(x, y)$ (see, [14,15]).

For many years, finite difference methods (cf. [16–22,9,23,24] and the reference therein) and finite element methods (cf. [25–29]) have been widely used to discrete the Helmholtz equation (1.1) with various boundaries. As is known to all, the solution of the Helmholtz equation oscillates severely for large wavenumbers, and the quality of the numerical results usually deteriorates as the wavenumber k increases (cf. [25–29]). Hence, there is a growing interest in discretization methods where the computational complexity increases only moderately with increasing wavenumber (cf. [25,28,30,20,22]). Compared with finite difference methods, discontinuous Galerkin (DG) methods can naturally handle inhomogeneous boundary conditions and curved boundaries, and they also allow the use of highly nonuniform and unstructured meshes. Moreover, DG methods also possess the property of higher accuracy. However, the finite difference method is more popular and useful in the numerical seismic wave propagation modeling (cf. [17,19]), because of its easier implementation and less computational complexity. Furthermore, we can easily extend finite difference methods to the three dimensional cases, and improve the accuracy of the solution of the Helmholtz equation by choosing optimal parameters in finite difference formulas [20,22].

In this paper, we are interested in the finite difference method because of its easy implementation, although its accuracy is usually lower than that of the finite element method. For accurate modeling, the conventional 5-point finite difference scheme requires 10 gridpoints per wavelength. Therefore, for the Helmholtz equation with large wavenumbers, the resulting matrix is very large and ill-conditioned. Usually, direct methods do not perform well, and iterative methods with preconditioners are alternative [31,17,24]. In 1996, the rotated 9-point finite difference scheme for the Helmholtz equation was proposed in [20] and one group of optimal parameters were also given. This optimal 9-point scheme reduces the number of gridpoints per wavelength to 5 while preserving the accuracy of the conventional 5-point scheme with 10 gridpoints per wavelength. Therefore, computer memory and CPU time are much saved. Shin and Sohn extended the idea of the rotated 9-point scheme to the 25-point formula in [22], which was called the extended 25-point finite difference scheme. The extended 25-point formula reduces the number of gridpoints per wavelength to 2, and its numerical dispersion is much less as compared with that of the rotated 9-point scheme.

Although it is good for the Helmholtz equation, the rotated 9-point scheme is not pointwise consistent with the Helmholtz-PML equation, which was proved in [13]. To overcome this difficulty, we presented a consistent 9-point finite difference scheme following the approach in [9]. Additionally, based on minimizing the numerical dispersion (see, [29,32]), global and refined choice strategies for choosing optimal parameters of the 9-point finite difference scheme were proposed. The improvement of the accuracy and the reduction of the numerical dispersion are significant. Moreover, to solve large scale problems, in [33] we proposed a new multigrid based preconditioned Krylov subspace method for the Helmholtz equation with PML, which was discretized by the optimal 9-point finite difference scheme proposed in [13]. The applications ranged from constant wavenumbers to irregular heterogeneity structures in a medium. The new multigrid based preconditioned Bi-CGSTAB method for the Helmholtz-PML equation was proved to be an efficient and robust iterative method. However, due to the existence of numerical dispersion, the numerical solution still deteriorates as the wavenumber k increases, although it has much improvement as compared with the rotated 9-point finite difference scheme. As the extended 25-point scheme has better properties than the rotated 9-point scheme (see, [22]), the goal of this paper is to propose an optimal 25-point scheme for the Helmholtz-PML equation, based on further minimizing the numerical dispersion. Moreover, we will find that, though the extended 25-point difference scheme is a popular solver for the Helmholtz equation, it is not a good choice for the Helmholtz equation with PML. We shall address this issue in detail in this paper.

This paper is organized as follows. In Section 2, we investigate the extended 25-point finite difference scheme and show that this scheme with the optimal parameters given in [22] is not pointwise consistent with the Helmholtz equation with PML. In Section 3, we present a 25-point difference scheme by using the approach suggested in paper [9], and prove that it is consistent with the Helmholtz equation with PML and is a second order scheme. We then propose the refined choice strategy for choosing optimal parameters of the 25-point finite difference scheme based on minimizing the numerical dispersion. In Section 4, numerical experiments are given to demonstrate the efficiency of the scheme. These numerical results show that the newly proposed schemes not only improve the accuracy, but also reduce the numerical dispersion significantly. Finally, Section 5 contains the conclusion of this paper.



(a) 0°. (b) 45°. (c) 0°. (d) 45°.

(e) θ_1. (f) θ_2.

Fig. 1. Five-point Laplacian operators.

2. The extended 25-point finite difference scheme for the Helmholtz-PML equation

In this section, we investigate the extended 25-point finite difference scheme for the Helmholtz equation with PML.

To construct finite difference schemes, we consider the network of grid points (x_m, y_n), where $x_m := x_0 + (m-1)h$ and $y_n := y_0 + (n-1)h$. Notice that the same step size $h := \Delta x = \Delta y$ is used for both variables x and y. Let $u_{m,n} := u\,|_{x=x_m,y=y_n}$ represent the pressure of the wave field at the location (x_m, y_n), and similarly, let $k_{m,n} := k\,|_{x=x_m,y=y_n}$ and $f_{m,n} := f\,|_{x=x_m,y=y_n}$. The key idea of the extended 25-point finite difference scheme is to approximate Δu by a second order centered difference using two 5-point 0° stars, two 45° rotated stars, the θ_1 rotated star and the θ_2 rotated star (see, Fig. 1), where $\theta_1 := \frac{180^\circ}{\pi}\arcsin\frac{\sqrt{5}}{5}$ and $\theta_2 := \frac{180^\circ}{\pi}\arcsin\frac{2\sqrt{5}}{5}$.

To develop the extended 25-point difference scheme for solving the Helmholtz-PML equation (1.2), we let

$$A_{m+\frac{i}{2},n+\frac{j}{2}} := A\left(x_0 + \left(m-1+\frac{i}{2}\right)\Delta x, y_0 + \left(n-1+\frac{j}{2}\right)\Delta y\right),$$

$$B_{m+\frac{i}{2},n+\frac{j}{2}} := B\left(x_0 + \left(m-1+\frac{i}{2}\right)\Delta x, y_0 + \left(n-1+\frac{j}{2}\right)\Delta y\right),$$

$$C_{m,n} := C(x_0 + (m-1)\Delta x, y_0 + (n-1)\Delta y),$$

for $i, j \in \mathbb{Z}_3 := \{-2, -1, 0, 1, 2\}$. According to the construction of the extended 25-point difference method, we define

$$\begin{aligned}\mathcal{L}_h u\,|_{x=x_m,y=y_n} := {} & a_1\mathcal{L}_{h,0^\circ}u\,|_{x=x_m,y=y_n} + a_2\mathcal{L}_{2h,0^\circ}u\,|_{x=x_m,y=y_n} + a_3\mathcal{L}_{h,45^\circ}u\,|_{x=x_m,y=y_n} + a_4\mathcal{L}_{2h,45^\circ}u\,|_{x=x_m,y=y_n} \\ & + a_5\mathcal{L}_{2h,\theta_1}u\,|_{x=x_m,y=y_n} + a_6\mathcal{L}_{2h,\theta_2}u\,|_{x=x_m,y=y_n},\end{aligned} \tag{2.1}$$

where $\sum_{j=1}^{6} a_j = 1$,

$$\begin{aligned}\mathcal{L}_{2h,\theta_2}u\,|_{x=x_m,y=y_n} := {} & \frac{1}{\sqrt{5}h}\left\{\left(A_{m+\frac{1}{2},n-1}\frac{u_{m+1,n-2}-u_{m,n}}{\sqrt{5}h} - A_{m-\frac{1}{2},n+1}\frac{u_{m,n}-u_{m-1,n+2}}{\sqrt{5}h}\right)\right. \\ & \left. + \left(B_{m+1,n+\frac{1}{2}}\frac{u_{m+2,n+1}-u_{m,n}}{\sqrt{5}h} - B_{m-1,n-\frac{1}{2}}\frac{u_{m,n}-u_{m-2,n-1}}{\sqrt{5}h}\right)\right\},\end{aligned} \tag{2.2}$$

and the former five terms on the left side of the Eq. (2.1) are constructed similarly based on the second order centered difference schemes for the point (x_m, y_n). We then approximate the first two terms of the left hand side of (1.2) as

$$\left[\frac{\partial}{\partial x}\left(A\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left(B\frac{\partial u}{\partial y}\right)\right]_{x=x_m,y=y_n} \approx \mathcal{L}_h u\,|_{x=x_m,y=y_n}.$$

Moreover, in order to approximate the term of zeroth order with 25 points, we let

$$\begin{aligned}
\mathcal{I}_{h,0^\circ}(Ck^2u)\,|_{x=x_m,y=y_n} &:= \frac{1}{4}\Big(C_{m-1,n}k^2_{m-1,n}u_{m-1,n} + C_{m+1,n}k^2_{m+1,n}u_{m+1,n} \\
&\quad + C_{m,n-1}k^2_{m,n-1}u_{m,n-1} + C_{m,n+1}k^2_{m,n+1}u_{m,n+1}\Big), \\
\mathcal{I}_{2h,0^\circ}(Ck^2u)\,|_{x=x_m,y=y_n} &:= \frac{1}{4}\Big(C_{m-2,n}k^2_{m-2,n}u_{m-2,n} + C_{m,n-2}k^2_{m,n-2}u_{m,n-2} \\
&\quad + C_{m+2,n}k^2_{m+2,n}u_{m+2,n} + C_{m,n+2}k^2_{m,n+2}u_{m,n+2}\Big), \\
\mathcal{I}_{h,45^\circ}(Ck^2u)\,|_{x=x_m,y=y_n} &:= \frac{1}{4}\Big(C_{m-1,n+1}k^2_{m-1,n+1}u_{m-1,n+1} + C_{m+1,n-1}k^2_{m+1,n-1}u_{m+1,n-1} \\
&\quad + C_{m-1,n-1}k^2_{m-1,n-1}u_{m-1,n-1} + C_{m+1,n+1}k^2_{m+1,n+1}u_{m+1,n+1}\Big), \\
\mathcal{I}_{2h,45^\circ}(Ck^2u)\,|_{x=x_m,y=y_n} &:= \frac{1}{4}\Big(C_{m-2,n-2}k^2_{m-2,n-2}u_{m-2,n-2} + C_{m+2,n-2}k^2_{m+2,n-2}u_{m+2,n-2} \\
&\quad + C_{m+2,n+2}k^2_{m+2,n+2}u_{m+2,n+2} + C_{m-2,n+2}k^2_{m-2,n+2}u_{m-2,n+2}\Big), \\
\mathcal{I}_{2h,\theta_1}(Ck^2u)\,|_{x=x_m,y=y_n} &:= \frac{1}{4}\Big(C_{m+2,n-1}k^2_{m+2,n-1}u_{m+2,n-1} + C_{m+1,n+2}k^2_{m+1,n+2}u_{m+1,n+2} \\
&\quad + C_{m-2,n+1}k^2_{m-2,n+1}u_{m-2,n+1} + C_{m-1,n-2}k^2_{m-1,n-2}u_{m-1,n-2}\Big), \\
\mathcal{I}_{2h,\theta_2}(Ck^2u)\,|_{x=x_m,y=y_n} &:= \frac{1}{4}\Big(C_{m+1,n-2}k^2_{m+1,n-2}u_{m+1,n-2} + C_{m+2,n+1}k^2_{m+2,n+1}u_{m+2,n+1} \\
&\quad + C_{m-1,n+2}k^2_{m-1,n+2}u_{m-1,n+2} + C_{m-2,n-1}k^2_{m-2,n-1}u_{m-2,n-1}\Big),
\end{aligned}$$

and approximate $Ck^2u\,|_{x=x_m,y=y_n}$ by a weighted average:

$$Ck^2u\,|_{x=x_m,y=y_n} \approx \mathcal{I}_h(Ck^2u)\,|_{x=x_m,y=y_n},$$

where

$$\begin{aligned}
\mathcal{I}_h(Ck^2u)\,|_{x=x_m,y=y_n} &:= b_1C_{m,n}k^2_{m,n}u_{m,n} + b_2\mathcal{I}_{h,0^\circ}(Ck^2u)\,|_{x=x_m,y=y_n} + b_3\mathcal{I}_{2h,0^\circ}(Ck^2u)\,|_{x=x_m,y=y_n} \\
&\quad + b_4\mathcal{I}_{h,45^\circ}(Ck^2u)\,|_{x=x_m,y=y_n} + b_5\mathcal{I}_{2h,45^\circ}(Ck^2u)\,|_{x=x_m,y=y_n} \\
&\quad + b_6\mathcal{I}_{2h,\theta_1}(Ck^2u)\,|_{x=x_m,y=y_n} + b_7\mathcal{I}_{2h,\theta_2}(Ck^2u)\,|_{x=x_m,y=y_n},
\end{aligned} \tag{2.3}$$

in which b_j $(j = 1, 2\ldots, 7)$ are parameters satisfying $\sum_{j=1}^{7} b_j = 1$. These yield the extended 25-point difference approximation for the Helmholtz-PML equation (1.2) as

$$\mathcal{L}_h u\,|_{x=x_m,y=y_n} + \mathcal{I}_h(Ck^2u)\,|_{x=x_m,y=y_n} = f_{m,n}. \tag{2.4}$$

In [22], a group of optimal parameters were given for the Helmholtz equation, that is,

$$\begin{aligned}
&a_1 = 0.0949098, \quad a_2 = 0.280677, \quad a_3 = 0.247253, \\
&a_4 = 0.0297441, \quad a_5 = 0.173708, \quad a_6 = 0.173708, \\
&b_1 = 0.363276, \quad b_2 = 0.434392, \quad b_3 = 0.0165948, \\
&b_4 = 0.1699204, \quad b_5 = 0.000825248, \quad b_6 = 0.0075106, \\
&b_7 = 0.00753368.
\end{aligned} \tag{2.5}$$

To analyze the relationship between the difference approximation (2.4) and the Helmholtz-PML equation (1.2), we recall the concept of consistency (cf. [34]).

Definition 2.1. Suppose that, the partial differential equation under consideration is $\mathcal{L}u = f$ and the corresponding finite difference approximation is $L_{m,n}U_{m,n} = F_{m,n}$, where $F_{m,n}$ denotes whatever approximation which has been made of the source term f. Let $(x_m, y_n) := (x_0 + (m-1)\Delta x, y_0 + (n-1)\Delta y)$. The finite difference scheme $L_{m,n}U_{m,n} = F_{m,n}$ is pointwise consistent with the partial differential equation $\mathcal{L}u = f$ at (x, y), if for any smooth function $\phi = \phi(x, y)$,

$$(\mathcal{L}\phi - f)\,|_{x=x_m,y=y_n} - \big[L_{m,n}\phi(x_m, y_n) - F_{m,n}\big] \to 0 \tag{2.6}$$

as $\Delta x, \Delta y \to 0$ and $(x_m, y_n) \to (x, y)$.

For the extended 25-point finite difference approximation (2.4), we have the following proposition.

Proposition 2.2. *The extended 25-point finite difference approximation (2.4) with optimal parameters (2.5) is not pointwise consistent with the Helmholtz-PML equation (1.2).*

Proof. Assume that, $x_m \le x < x_{m+1}$ and $y_n \le y < y_{n+1}$. It follows from the Taylor theorem that

$$\mathcal{L}_{h,0}\circ u\,|_{x=x_m,y=y_n} = \frac{\partial}{\partial x}\left(A\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left(B\frac{\partial u}{\partial y}\right) + \mu_1 h^2 + \mathcal{O}(h^3), \tag{2.7}$$

$$\mathcal{L}_{2h,0}\circ u\,|_{x=x_m,y=y_n} = \frac{\partial}{\partial x}\left(A\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left(B\frac{\partial u}{\partial y}\right) + 4\mu_1 h^2 + \mathcal{O}(h^3), \tag{2.8}$$

$$\mathcal{L}_{h,45}\circ u\,|_{x=x_m,y=y_n} = E + \mu_2 h^2 + \mathcal{O}(h^3), \tag{2.9}$$

$$\mathcal{L}_{2h,45}\circ u\,|_{x=x_m,y=y_n} = E + 4\mu_2 h^2 + \mathcal{O}(h^3), \tag{2.10}$$

$$\begin{aligned}\mathcal{L}_{2h,\theta_1} u\,|_{x=x_m,y=y_n} \;=\; & \frac{1}{5}\left\{\frac{\partial}{\partial x}\left((4A+B)\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left((A+4B)\frac{\partial u}{\partial y}\right) + 2\frac{\partial}{\partial x}\left((B-A)\frac{\partial u}{\partial y}\right)\right.\\ & \left. + 2\frac{\partial}{\partial y}\left((B-A)\frac{\partial u}{\partial x}\right)\right\} + \mu_3 h^2 + \mathcal{O}(h^3),\end{aligned} \tag{2.11}$$

$$\begin{aligned}\mathcal{L}_{2h,\theta_2} u\,|_{x=x_m,y=y_n} \;=\; & \frac{1}{5}\left\{\frac{\partial}{\partial x}\left((A+4B)\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left((4A+B)\frac{\partial u}{\partial y}\right) + 2\frac{\partial}{\partial x}\left((B-A)\frac{\partial u}{\partial y}\right)\right.\\ & \left. + 2\frac{\partial}{\partial y}\left((B-A)\frac{\partial u}{\partial x}\right)\right\} + \mu_4 h^2 + \mathcal{O}(h^3),\end{aligned} \tag{2.12}$$

in which

$$\mu_1 := \frac{1}{24}\left\{\frac{\partial^3}{\partial x^3}\left(A\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial x}\left(A\frac{\partial^3 u}{\partial x^3}\right) + \frac{\partial^3}{\partial y^3}\left(B\frac{\partial u}{\partial y}\right) + \frac{\partial}{\partial y}\left(B\frac{\partial^3 u}{\partial y^3}\right)\right\},$$

$$E := \frac{1}{2}\left\{\frac{\partial}{\partial x}\left((A+B)\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left((A+B)\frac{\partial u}{\partial y}\right) + \frac{\partial}{\partial x}\left((B-A)\frac{\partial u}{\partial y}\right) + \frac{\partial}{\partial y}\left((B-A)\frac{\partial u}{\partial x}\right)\right\},$$

$$\begin{aligned}\mu_2 := \; & \frac{1}{48}\left\{\left(\frac{\partial}{\partial x}-\frac{\partial}{\partial y}\right)^3\left(A\left(\frac{\partial}{\partial x}-\frac{\partial}{\partial y}\right)u\right) + \left(\frac{\partial}{\partial x}-\frac{\partial}{\partial y}\right)\left(A\left(\frac{\partial}{\partial x}-\frac{\partial}{\partial y}\right)^3 u\right)\right.\\ & \left. + \left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)^3\left(B\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)u\right) + \left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)\left(B\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)^3 u\right)\right\},\end{aligned}$$

$$\begin{aligned}\mu_3 := \; & \frac{1}{15}\left\{\left(\frac{\partial}{\partial x}-\frac{1}{2}\frac{\partial}{\partial y}\right)^3\left(A\left(2\frac{\partial}{\partial x}-\frac{\partial}{\partial y}\right)u\right) + \left(2\frac{\partial}{\partial x}-\frac{\partial}{\partial y}\right)\left(A\left(\frac{\partial}{\partial x}-\frac{1}{2}\frac{\partial}{\partial y}\right)^3 u\right)\right.\\ & \left. + \left(\frac{1}{2}\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)^3\left(B\left(\frac{\partial}{\partial x}+2\frac{\partial}{\partial y}\right)u\right) + \left(\frac{\partial}{\partial x}+2\frac{\partial}{\partial y}\right)\left(B\left(\frac{1}{2}\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)^3 u\right)\right\},\end{aligned}$$

$$\begin{aligned}\mu_4 := \; & \frac{1}{15}\left\{\left(\frac{1}{2}\frac{\partial}{\partial x}-\frac{\partial}{\partial y}\right)^3\left(A\left(\frac{\partial}{\partial x}-2\frac{\partial}{\partial y}\right)u\right) + \left(\frac{\partial}{\partial x}-2\frac{\partial}{\partial y}\right)\left(A\left(\frac{1}{2}\frac{\partial}{\partial x}-\frac{\partial}{\partial y}\right)^3 u\right)\right.\\ & \left. + \left(\frac{\partial}{\partial x}+\frac{1}{2}\frac{\partial}{\partial y}\right)^3\left(B\left(2\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)u\right) + \left(2\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)\left(B\left(\frac{\partial}{\partial x}+\frac{1}{2}\frac{\partial}{\partial y}\right)^3 u\right)\right\}.\end{aligned}$$

Similarly, we have

$$\mathcal{I}_h(Ck^2 u) = Ck^2 u + \mu_5 h^2 + \mathcal{O}(h^3), \tag{2.13}$$

in which

$$\mu_5 = \left(\frac{b_2}{4} + b_3 + \frac{b_4}{2} + 2b_5 + \frac{5b_6}{4} + \frac{5b_7}{4}\right)\left(\frac{\partial^2}{\partial x^2}(Ck^2 u) + \frac{\partial^2}{\partial y^2}(Ck^2 u)\right).$$

It follows from equations (2.7)–(2.13) that the left hand side of the extended 25-point finite difference approximation (2.4) is equivalent to

$$\begin{aligned}&\left(a_1+a_2+\frac{a_3+a_4}{2}+\frac{4}{5}a_5+\frac{1}{5}a_6\right)\left\{\frac{\partial}{\partial x}\left(A\frac{\partial u}{\partial x}\right)+\frac{\partial}{\partial y}\left(B\frac{\partial u}{\partial y}\right)\right\}\\&+\left(\frac{a_3+a_4}{2}+\frac{2}{5}a_5+\frac{2}{5}a_6\right)\left\{\frac{\partial}{\partial x}\left((B-A)\frac{\partial u}{\partial y}\right)+\frac{\partial}{\partial y}\left((B-A)\frac{\partial u}{\partial x}\right)\right\}\\&+\left(\frac{a_3+a_4}{2}+\frac{1}{5}a_5+\frac{4}{5}a_6\right)\left\{\frac{\partial}{\partial y}\left(A\frac{\partial u}{\partial y}\right)+\frac{\partial}{\partial x}\left(B\frac{\partial u}{\partial x}\right)\right\}+\zeta_{h^2}h^2+\mathcal{O}(h^3),\end{aligned}$$

in which

$$\zeta_{h^2}=(a_1+4a_2)\mu_1+(a_3+4a_4)\mu_2+a_5\mu_3+a_6\mu_4+\mu_5.$$

Since $A\neq B$ almost everywhere in PML, the necessary condition for the consistency is

$$\begin{cases}a_1+a_2+\dfrac{a_3+a_4}{2}+\dfrac{4}{5}a_5+\dfrac{1}{5}a_6=1,\\[2mm]\dfrac{a_3+a_4}{2}+\dfrac{2}{5}a_5+\dfrac{2}{5}a_6=0,\\[2mm]\dfrac{a_3+a_4}{2}+\dfrac{1}{5}a_5+\dfrac{4}{5}a_6=0.\end{cases}$$

The above condition is not satisfied by the group of optimal parameters (2.5). Therefore, we come to the conclusion of this proposition. □

From this proposition we conclude that, the extended 25-point scheme with optimal parameters (2.5) is not a good choice for the Helmholtz-PML equation, thus we need to adapt the difference scheme to our problem.

3. An optimal 25-point finite difference scheme for the Helmholtz equation with PML

In this section, we propose a 25-point finite difference scheme which is pointwise consistent with the Helmholtz-PML equation, and then present a refined optimization rule for choosing the parameters of the finite difference scheme such that the numerical dispersion is minimized well. Finally, we generalize the scheme to a case with non-equidistant grid.

3.1. A consistent 25-point difference scheme

First, we propose a consistent 25-point finite difference scheme for the Helmholtz-PML equation. The approach of constructing this kind of difference schemes was proposed in [9] and further developed in [13].

We denote

$$\tilde{\mathcal{L}}_{h,x}u\,|_{(m,n+j)}:=\frac{A_{m+\frac{1}{2},n+j}(u_{m+1,n+j}-u_{m,n+j})-A_{m-\frac{1}{2},n+j}(u_{m,n+j}-u_{m-1,n+j})}{h^2},\tag{3.1}$$

for $j\in\mathbb{Z}_2:=\{-1,0,1\}$,

$$\tilde{\mathcal{L}}_{2h,x}u\,|_{(m,n+j)}:=\frac{A_{m+1,n+j}(u_{m+2,n+j}-u_{m,n+j})-A_{m-1,n+j}(u_{m,n+j}-u_{m-2,n+j})}{(2h)^2},\tag{3.2}$$

for $j\in\mathbb{Z}_3$, and let

$$\begin{aligned}\tilde{\mathcal{L}}_{h,x}u\,|_{x=x_m,y=y_n}&:=\gamma_1\tilde{\mathcal{L}}_{h,x}u\,|_{(m,n)}+\frac{\gamma_2}{2}\left(\tilde{\mathcal{L}}_{h,x}u\,|_{(m,n-1)}+\tilde{\mathcal{L}}_{h,x}u\,|_{(m,n+1)}\right),\\\tilde{\mathcal{L}}_{2h,x}u\,|_{x=x_m,y=y_n}&:=\gamma_3\tilde{\mathcal{L}}_{2h,x}u\,|_{(m,n)}+\frac{\gamma_4}{2}\left(\tilde{\mathcal{L}}_{2h,x}u\,|_{(m,n-2)}+\tilde{\mathcal{L}}_{2h,x}u\,|_{(m,n+2)}\right)\\&\quad+\frac{\gamma_5}{2}\left(\tilde{\mathcal{L}}_{2h,x}u\,|_{(m,n-1)}+\tilde{\mathcal{L}}_{2h,x}u\,|_{(m,n+1)}\right),\end{aligned}$$

where $\sum_{j=1}^{5}\gamma_j=1$. Then we approximate the first term of the left hand side of (1.2) as

$$\frac{\partial}{\partial x}\left(A\frac{\partial u}{\partial x}\right)|_{x=x_m,y=y_n}\approx\tilde{\mathcal{L}}_{h,x}u\,|_{x=x_m,y=y_n}+\tilde{\mathcal{L}}_{2h,x}u\,|_{x=x_m,y=y_n}.$$

We deal with the approximation of the second term in a similar way, that is,

$$\frac{\partial}{\partial y}\left(B\frac{\partial u}{\partial y}\right)|_{x=x_m,y=y_n}\approx\tilde{\mathcal{L}}_{h,y}u\,|_{x=x_m,y=y_n}+\tilde{\mathcal{L}}_{2h,y}u\,|_{x=x_m,y=y_n}.$$

Let $\tilde{\mathcal{L}}_h := \tilde{\mathcal{L}}_{h,x}+\tilde{\mathcal{L}}_{h,y}+\tilde{\mathcal{L}}_{2h,x}+\tilde{\mathcal{L}}_{2h,y}$. We obtain the following 25-point finite difference approximation for the Helmholtz-PML equation (1.2)

$$\tilde{\mathcal{L}}_h u\,|_{x=x_m,y=y_n} + \mathcal{I}_h(Ck^2u)\,|_{x=x_m,y=y_n} = f_{m,n}. \tag{3.3}$$

The above finite difference scheme tells us how to discrete the interior points. Next, we give some remarks on how to deal with boundary points.

Remark 3.1. As the PML has a good property of absorbing outgoing waves, we assume that there are no waves outside the interior domain and PML. Computationally, we deal with the boundary points (x_m, y_n) in the same way as interior points except that, we simply set $u_{m+i,n+j} = 0$ if the point (x_{m+i}, y_{n+j}) lies outside PML.

The next proposition presents the error analysis for the 25-point difference scheme (3.3).

Proposition 3.2. *If $\sum_{j=1}^{5}\gamma_j = 1$ and $\sum_{j=1}^{7} b_j = 1$, then 25-point finite difference approximation (3.3) is pointwise consistent with the Helmholtz-PML equation (1.2) and is a second order scheme.*

Proof. Assume that $x_m \le x < x_{m+1}$ and $y_n \le y < y_{n+1}$. It follows from the Taylor theorem that

$$\tilde{\mathcal{L}}_{h,x}u\,|_{x=x_m,y=y_n} = (\gamma_1+\gamma_2)\frac{\partial}{\partial x}\left(A\frac{\partial u}{\partial x}\right) + \nu_1 h^2 + \mathcal{O}(h^3), \tag{3.4}$$

$$\tilde{\mathcal{L}}_{h,y}u\,|_{x=x_m,y=y_n} = (\gamma_1+\gamma_2)\frac{\partial}{\partial y}\left(B\frac{\partial u}{\partial y}\right) + \nu_2 h^2 + \mathcal{O}(h^3), \tag{3.5}$$

in which

$$\nu_1 := \frac{1}{24}(\gamma_1+\gamma_2)\left\{\frac{\partial^3}{\partial x^3}\left(A\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial x}\left(A\frac{\partial^3 u}{\partial x^3}\right)\right\} + \frac{\gamma_2}{2}\frac{\partial^3}{\partial y^2\partial x}\left(A\frac{\partial u}{\partial x}\right),$$

$$\nu_2 := \frac{1}{24}(\gamma_1+\gamma_2)\left\{\frac{\partial^3}{\partial y^3}\left(B\frac{\partial u}{\partial y}\right) + \frac{\partial}{\partial y}\left(B\frac{\partial^3 u}{\partial y^3}\right)\right\} + \frac{\gamma_2}{2}\frac{\partial^3}{\partial x^2\partial y}\left(B\frac{\partial u}{\partial y}\right).$$

Similarly, we have

$$\tilde{\mathcal{L}}_{2h,x}u\,|_{x=x_m,y=y_n} = (\gamma_3+\gamma_4+\gamma_5)\frac{\partial}{\partial x}\left(A\frac{\partial u}{\partial x}\right) + \nu_3 h^2 + \mathcal{O}(h^3), \tag{3.6}$$

$$\tilde{\mathcal{L}}_{2h,y}u\,|_{x=x_m,y=y_n} = (\gamma_3+\gamma_4+\gamma_5)\frac{\partial}{\partial y}\left(B\frac{\partial u}{\partial y}\right) + \nu_4 h^2 + \mathcal{O}(h^3), \tag{3.7}$$

where

$$\begin{aligned}\nu_3 :=\ & \gamma_3\left\{\frac{1}{12}\frac{\partial^3}{\partial x^3}\left(A\frac{\partial u}{\partial x}\right) + \frac{1}{3}\frac{\partial}{\partial x}\left(A\frac{\partial^3 u}{\partial x^3}\right)\right\} + \frac{\gamma_4}{2}\left\{4\frac{\partial^3}{\partial y^2\partial x}\left(A\frac{\partial u}{\partial x}\right) + \frac{1}{6}\frac{\partial^3}{\partial x^3}\left(A\frac{\partial u}{\partial x}\right) + \frac{2}{3}\frac{\partial}{\partial x}\left(A\frac{\partial^3 u}{\partial x^3}\right)\right\}\\ &+ \frac{\gamma_5}{2}\left\{\frac{\partial^3}{\partial y^2\partial x}\left(A\frac{\partial u}{\partial x}\right) + \frac{1}{6}\frac{\partial^3}{\partial x^3}\left(A\frac{\partial u}{\partial x}\right) + \frac{2}{3}\frac{\partial}{\partial x}\left(A\frac{\partial^3 u}{\partial x^3}\right)\right\},\\ \nu_4 :=\ & \gamma_3\left\{\frac{1}{12}\frac{\partial^3}{\partial y^3}\left(B\frac{\partial u}{\partial y}\right) + \frac{1}{3}\frac{\partial}{\partial y}\left(B\frac{\partial^3 u}{\partial y^3}\right)\right\} + \frac{\gamma_4}{2}\left\{4\frac{\partial^3}{\partial x^2\partial y}\left(B\frac{\partial u}{\partial y}\right) + \frac{1}{6}\frac{\partial^3}{\partial y^3}\left(B\frac{\partial u}{\partial y}\right) + \frac{2}{3}\frac{\partial}{\partial y}\left(B\frac{\partial^3 u}{\partial y^3}\right)\right\}\\ &+ \frac{\gamma_5}{2}\left\{\frac{\partial^3}{\partial x^2\partial y}\left(B\frac{\partial u}{\partial y}\right) + \frac{1}{6}\frac{\partial^3}{\partial y^3}\left(B\frac{\partial u}{\partial y}\right) + \frac{2}{3}\frac{\partial}{\partial y}\left(B\frac{\partial^3 u}{\partial y^3}\right)\right\}.\end{aligned}$$

It follows from (3.4)–(3.7) and (2.13) that the left hand side of the 25-point finite difference approximation (3.3) is equivalent to

$$\frac{\partial}{\partial x}\left(A\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left(B\frac{\partial u}{\partial y}\right) + Ck^2u + \theta_{h^2}h^2 + \mathcal{O}(h^3), \tag{3.8}$$

where $\theta_{h^2} = \sum_{j=1}^{4}\nu_j + \mu_5$. From (1.2) and (3.8) we conclude the results of this proposition. □

From the proposition above, we see that the 25-point finite difference scheme (3.3) is a second order scheme for arbitrary constants γ_j ($j \in \mathbb{N}_5 := \{1, 2, \ldots, 5\}$) and b_j ($j \in \mathbb{N}_7 := \{1, 2, \ldots, 7\}$) provided $\sum_{j=1}^{5}\gamma_j = 1$ and $\sum_{j=1}^{7} b_j = 1$. A further observation yields the following proposition.

Proposition 3.3. *The 25-point difference schemes (2.4) and (3.3) are equivalent in the interior area if*

$$\begin{aligned} &a_5 = a_6, \qquad \gamma_5 = \frac{8}{5}a_5, \qquad \gamma_4 = \frac{a_4}{2}, \\ &\gamma_3 = a_2 + \frac{a_4}{2}, \qquad \gamma_2 = \frac{a_3}{2}, \qquad \gamma_1 = a_1 + \frac{1}{2}a_3 + \frac{2}{5}a_5. \end{aligned} \tag{3.9}$$

Proof. For brevity, we only give the proof under the condition that k is a positive constant. When k is a variable, the same conclusion can be obtained similarly.

In the interior area, $A = B = C = 1$, thus replacing $u_{m+i,n+j}$ with $U_{m+i,n+j}$ $(i, j \in \mathbb{Z}_3)$ in the formula (3.3) gives

$$\begin{aligned} &T_1U_{m-2,n-2} + T_2U_{m-1,n-2} + T_4U_{m,n-2} + T_3U_{m+1,n-2} + T_1U_{m+2,n-2} \\ &\quad + T_3U_{m-2,n-1} + T_5U_{m-1,n-1} + T_6U_{m,n-1} + T_5U_{m+1,n-2} + T_2U_{m+2,n-1} \\ &\quad + T_4U_{m-2,n} + T_6U_{m-1,n} + T_0U_{m,n} + T_6U_{m+1,n} + T_4U_{m+2,n} \\ &\quad + T_2U_{m-2,n+1} + T_5U_{m-1,n+1} + T_6U_{m,n+1} + T_5U_{m+1,n+1} + T_3U_{m+2,n+1} \\ &\quad + T_1U_{m-2,n+2} + T_3U_{m-1,n+2} + T_4U_{m,n+2} + T_2U_{m+1,n+2} + T_1U_{m+2,n+2} \\ &= g_{m,n}, \end{aligned} \tag{3.10}$$

in which

$$\begin{aligned} &T_0 := -\frac{1}{h^2}(4\gamma_1 + \gamma_3) + b_1k^2, \qquad T_1 := \frac{\gamma_4}{4h^2} + \frac{b_5}{4}k^2, \qquad T_2 := \frac{\gamma_5}{8h^2} + \frac{b_6}{4}k^2, \\ &T_3 := \frac{\gamma_5}{8h^2} + \frac{b_7}{4}k^2, \qquad T_4 := \frac{1}{4h^2}(\gamma_3 - \gamma_4) + \frac{b_3}{4}k^2, \qquad T_5 := \frac{\gamma_2}{h^2} + \frac{b_4}{4}k^2, \\ &T_6 := \frac{1}{h^2}\left(-\gamma_2 + \gamma_1 - \frac{\gamma_5}{4}\right) + \frac{b_2}{4}k^2. \end{aligned}$$

Then replacing $u_{m,n}$ with its finite difference approximation $U_{m,n}$ in the extended 25-point finite difference approximation, (2.4) yields

$$\begin{aligned} &\hat{T}_1U_{m-2,n-2} + \hat{T}_2U_{m-1,n-2} + \hat{T}_4U_{m,n-2} + \hat{T}_3U_{m+1,n-2} + \hat{T}_1U_{m+2,n-2} \\ &\quad + \hat{T}_3U_{m-2,n-1} + \hat{T}_5U_{m-1,n-1} + \hat{T}_6U_{m,n-1} + \hat{T}_5U_{m+1,n-2} + \hat{T}_2U_{m+2,n-1} \\ &\quad + \hat{T}_4U_{m-2,n} + \hat{T}_6U_{m-1,n} + \hat{T}_0U_{m,n} + \hat{T}_6U_{m+1,n} + \hat{T}_4U_{m+2,n} \\ &\quad + \hat{T}_2U_{m-2,n+1} + \hat{T}_5U_{m-1,n+1} + \hat{T}_6U_{m,n+1} + \hat{T}_5U_{m+1,n+1} + \hat{T}_3U_{m+2,n+1} \\ &\quad + \hat{T}_1U_{m-2,n+2} + \hat{T}_3U_{m-1,n+2} + \hat{T}_4U_{m,n+2} + \hat{T}_2U_{m+1,n+2} + \hat{T}_1U_{m+2,n+2} \\ &= g_{m,n}, \end{aligned} \tag{3.11}$$

where

$$\begin{aligned} &\hat{T}_0 := \frac{1}{h^2}\left(-4a_1 - a_2 - 2a_3 - \frac{a_4}{2} - \frac{4}{5}a_5 - \frac{4}{5}a_6\right) + b_1k^2, \\ &\hat{T}_1 := \frac{a_4}{8h^2} + \frac{b_5}{4}k^2, \qquad \hat{T}_2 := \frac{a_5}{5h^2} + \frac{b_6}{4}k^2, \qquad \hat{T}_3 := \frac{a_6}{5h^2} + \frac{b_7}{4}k^2, \\ &\hat{T}_4 := \frac{a_2}{4h^2} + \frac{b_3}{4}k^2, \qquad \hat{T}_5 := \frac{a_3}{2h^2} + \frac{b_4}{4}k^2, \qquad \hat{T}_6 := \frac{a_1}{h^2} + \frac{b_2}{4}k^2. \end{aligned}$$

Comparing the parameters of (3.10) and (3.11) leads to the result of this proposition. □

From Proposition 3.3, in the interior area, we have the extended 25-point finite difference scheme (2.4) with optimal parameters (2.5) equivalent to the 25-point finite difference scheme (3.3) with the parameters below

$$\begin{aligned} &\gamma_1 = 0.2880195, \qquad \gamma_2 = 0.12362650, \qquad \gamma_3 = 0.29554905, \\ &\gamma_4 = 0.014872050, \qquad \gamma_5 = 0.27793280, \qquad b_1 = 0.363276, \\ &b_2 = 0.434392, \qquad b_3 = 0.0165948, \qquad b_4 = 0.1699204, \\ &b_5 = 0.000825248, \qquad b_6 = 0.0075106, \qquad b_7 = 0.00753368. \end{aligned} \tag{3.12}$$

In order to distinguish with the extended 25-point scheme (the extended 25p), we call the difference scheme (3.3) with parameters (3.12) as the global optimal 25-point scheme for the Helmholtz-PML equation (or simply the global 25p).

3.2. Choice strategies for optimal parameters of the finite difference scheme

We now go further into the problem of the choice strategies for optimal parameters. Since the solution of the Helmholtz equation is oscillating seriously for large wavenumbers, for a finite difference scheme, only the convergence order is not

enough. In fact, the accuracy of the numerical solution deteriorates with increasing wavenumber k. The phenomenon is the so-called 'pollution effect'. As the result of the 'pollution', the wavenumber of the numerical solution is different from the wavenumber of the exact solution, and this is the so called 'numerical dispersion' (see, [29,32]). Therefore, in order to reduce the 'numerical dispersion', we need to minimize the error between the numerical wavenumber and the exact wavenumber. If the difference scheme has optimal convergence order, and the parameters are chosen such that the scheme has minimal numerical dispersion in the interior area, then we regard it as an optimal scheme for the Helmholtz-PML equation (see, [13,21]).

To optimize the 25-point scheme (3.3), we first assume a plane-wave solution of the form $U(x, y) = \mathrm{e}^{-\mathrm{i}k(x\cos\theta+y\sin\theta)}$, where θ is the propagation angle from the y-axis. Moreover, the classical dispersion analysis is generally done in the infinite domain with the source $g = 0$ and the wavenumber k being a positive constant (see, [20,22]). For the convenience of analysis, let v be the velocity of propagation, λ be the wavelength, and G be the number of gridpoints per wavelength, that is, $G = \frac{\lambda}{h}$. Since $\lambda = \frac{2\pi v}{\omega}$ and $k = \frac{\omega}{v}$, we have $kh = \frac{2\pi}{G}$. Also, denote

$$P_1 := \cos(kh\sin\theta) = \cos\left(\frac{2\pi}{G}\sin\theta\right), \qquad Q_1 := \cos(kh\cos\theta) = \cos\left(\frac{2\pi}{G}\cos\theta\right),$$

$$P_2 := \cos(2kh\sin\theta) = \cos\left(\frac{4\pi}{G}\sin\theta\right), \qquad Q_2 := \cos(2kh\cos\theta) = \cos\left(\frac{4\pi}{G}\cos\theta\right),$$

$$R_1 := \cos(kh(2\sin\theta - \cos\theta)) = \cos\left(\frac{2\pi}{G}(2\sin\theta - \cos\theta)\right),$$

$$R_2 := \cos(kh(\sin\theta + 2\cos\theta)) = \cos\left(\frac{2\pi}{G}(\sin\theta + 2\cos\theta)\right),$$

$$R_3 := \cos(kh(\sin\theta - 2\cos\theta)) = \cos\left(\frac{2\pi}{G}(\sin\theta - 2\cos\theta)\right),$$

$$R_4 := \cos(kh(2\sin\theta + \cos\theta)) = \cos\left(\frac{2\pi}{G}(2\sin\theta + \cos\theta)\right).$$

By substituting $U_{m+j,n+l} := \mathrm{e}^{-\mathrm{i}k(x_{m+j}\cos\theta+y_{n+l}\sin\theta)}$ $(j, l \in \mathbb{Z}_3)$ into Eq. (3.10), dividing both sides by the factor $\mathrm{e}^{-\mathrm{i}k(x_m\cos\theta+y_n\sin\theta)}$, and finally applying the Euler formula $\mathrm{e}^{\mathrm{i}x} = \cos x + \mathrm{i}\sin x$ lead to the dispersion equation

$$k^2h^2L = R, \tag{3.13}$$

where

$$L := b_1 + \frac{b_2}{2}(P_1 + Q_1) + \frac{b_3}{2}(P_2 + Q_2) + b_4P_1Q_1 + b_5P_2Q_2 + \frac{b_6}{2}(R_1 + R_2) + \frac{b_7}{2}(R_3 + R_4),$$

$$\begin{aligned} R := {} & 2\gamma_1(2 - P_1 - Q_1) + 2\gamma_2(P_1 + Q_1 - 2P_1Q_1) + \frac{\gamma_3}{2}(2 - P_2 - Q_2) \\ & + \frac{\gamma_4}{2}(P_2 + Q_2 - 2P_2Q_2) + \frac{\gamma_5}{2}(Q_1 - P_2Q_1 + P_1 - P_1Q_2). \end{aligned}$$

Then replacing k on the left side of the Eq. (3.13) with the numerical wavenumber k^N yields

$$k^N = \frac{1}{h}\sqrt{\frac{R}{L}}. \tag{3.14}$$

With the Taylor expansion, we have

$$\begin{aligned} (k^N)^2 = {} & k^2 + \left\{-\frac{1}{12}[\gamma_1 + \gamma_2 + 4(\gamma_3 + \gamma_4 + \gamma_5)] + \left[\frac{b_2}{4} + b_3 + \frac{b_4}{2} + 2b_5 + \frac{5}{4}(b_6 + b_7)\right]\right. \\ & \left. + \frac{1}{24}(\gamma_1 - 5\gamma_2 + 4\gamma_3 - 20\gamma_4 - 2\gamma_5)\sin^2(2\theta)\right\}k^4h^2 + \mathcal{O}(k^5h^3), \quad kh \to 0. \end{aligned} \tag{3.15}$$

The above equation indicates that $(k^N)^2$ approximates k^2 in a second order. Moreover, the term associated with k^4h^2 presents the pollution effect, which depends on the wavenumber k, the parameters of the finite difference formula (3.3) and the wave's propagation angle θ from the y-axis. For the optimal 9-point finite difference scheme proposed in [13], a similar estimate for the relation $(k^N)^2$ and k^2 is obtained; see Remark 3.4. It is easy to see that, for both the 25-point scheme (3.3) and the 9-point scheme proposed in [13], the numerical wavenumber k^N approximates the exact wavenumber k in the same order. However, with more parameters in the difference formula (3.3), we can better suppress the pollution effect.

Remark 3.4. For the optimal 9-point finite difference scheme proposed in [13], we have

$$(k^N)^2 = k^2 + \left[\frac{d}{4} + \frac{e}{2} - \frac{1}{12} + \left(\frac{b}{4} - \frac{5}{24}\right)\sin^2(2\theta)\right]k^4h^2 + \mathcal{O}(k^5h^3), \quad kh \to 0, \tag{3.16}$$

where b, d, e are parameters of the optimal 9-point formula.

Next, we present the relation of the numerical wavenumber k^N and the exact wavenumber k. Since $h = \frac{2\pi}{Gk}$, we conclude that

$$\frac{k^N}{k} = \frac{G}{2\pi}\sqrt{\frac{R}{L}}. \tag{3.17}$$

Moreover, the normalized numerical phase velocity (cf. [20,22,35]) is

$$\frac{V_{ph}^N}{v} = \frac{G}{2\pi}\sqrt{\frac{R}{L}}, \tag{3.18}$$

and the normalized numerical group velocity is

$$\frac{V_{gr}^N}{v} = \frac{G}{4\pi}\frac{v}{V_{ph}^N}\left\{\frac{\left(\frac{1}{h}\frac{\partial R}{\partial k}\right)L - R\left(\frac{1}{h}\frac{\partial L}{\partial k}\right)}{L^2}\right\}. \tag{3.19}$$

We choose optimal parameters γ_i ($i \in \mathbb{N}_5$) and b_j ($j \in \mathbb{N}_7$) by minimizing the numerical dispersion. To do this, we set

$$J(\gamma_1, \ldots, \gamma_5, b_1, \ldots, b_7; G, \theta) := \frac{G}{2\pi}\sqrt{\frac{R}{L}} - 1, \tag{3.20}$$

for $\sum_{j=1}^5 \gamma_j = 1$, $\sum_{j=1}^7 b_j = 1$ and $(G, \theta) \in I_G \times I_\theta$, where I_G and I_θ are two intervals. In general, one can choose $I_\theta := \left[0, \frac{\pi}{2}\right]$ and $I_G := [G_{\min}, G_{\max}] = [4, 400]$. We remark that the interval $\left[0, \frac{\pi}{2}\right]$ can be replaced by $\left[0, \frac{\pi}{4}\right]$ because of the symmetry, and $G_{\min} \geq 2$ based on the Nyquist sampling limit (cf. [22]).

It follows from (3.17) that minimizing the error between the numerical wavenumber k^N and the exact wavenumber k is equivalent to minimizing the norm $\|J(\gamma_1, \ldots, \gamma_5, b_1, \ldots, b_7; \cdot, \cdot)\|_{\infty, I_G \times I_\theta}$. To do this, we first set $J(\gamma_1, \ldots, \gamma_5, b_1, \ldots, b_7; G, \theta) = 0$, which yields the equation

$$\frac{G^2}{4\pi^2}\frac{R}{L} = 1.$$

Thus, we have that

$$\begin{aligned}
&G^2\left\{4\left(P_1Q_1 - Q_1 - P_1 + 1\right)\gamma_2 + \left[3 - 2(P_1 + Q_1) + \frac{1}{2}\left(P_2 + Q_2\right)\right]\gamma_3\right.\\
&\left. + \frac{1}{2}[2P_2Q_2 - 4(P_1 + Q_1) - P_2 - Q_2 + 8]\gamma_4 + \frac{1}{2}[P_1Q_2 + P_2Q_1 - 5(P_1 + Q_1) + 8]\gamma_5\right\}\\
&+ 4\pi^2\left\{\frac{1}{2}(P_1 + Q_1 - 2)b_2 + \frac{1}{2}(P_2 + Q_2 - 2)b_3 + (P_1Q_1 - 1)b_4 + (P_2Q_2 - 1)b_5\right.\\
&\left. + \frac{1}{2}\left(R_1 + R_2 - 2\right)b_6 + \frac{1}{2}\left(R_3 + R_4 - 2\right)b_7\right\} = -4\pi^2 + 2G^2(2 - P_1 - Q_1).
\end{aligned} \tag{3.21}$$

Note that P_1, P_2, Q_1, Q_2, R_1, R_2, R_3 and R_4 are functions of G and θ. We choose $\theta = \theta_m := \frac{(m-1)\pi}{4(l-1)} \in I_\theta$, $m = 1, 2, \ldots, l$, and $\frac{1}{G} = \frac{1}{G_n} := \frac{1}{G_{\max}} + (n-1)\frac{\frac{1}{G_{\min}} - \frac{1}{G_{\max}}}{r-1} \in \left[\frac{1}{G_{\max}}, \frac{1}{G_{\min}}\right]$, $n = 1, 2, \ldots, r$. Then, Eq. (3.21) leads to the linear system

$$\begin{bmatrix}
S_{1,1}^1 & S_{1,1}^2 & S_{1,1}^3 & S_{1,1}^4 & S_{1,1}^5 & S_{1,1}^6 & S_{1,1}^7 & S_{1,1}^8 & S_{1,1}^9 & S_{1,1}^{10}\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
S_{1,r}^1 & S_{1,r}^2 & S_{1,r}^3 & S_{1,r}^4 & S_{1,r}^5 & S_{1,r}^6 & S_{1,r}^7 & S_{1,r}^8 & S_{1,r}^9 & S_{1,r}^{10}\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
S_{m,n}^1 & S_{m,n}^2 & S_{m,n}^3 & S_{m,n}^4 & S_{m,n}^5 & S_{m,n}^6 & S_{m,n}^7 & S_{m,n}^8 & S_{m,n}^9 & S_{m,n}^{10}\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
S_{l,r}^1 & S_{l,r}^2 & S_{l,r}^3 & S_{l,r}^4 & S_{l,r}^5 & S_{l,r}^6 & S_{l,r}^7 & S_{l,r}^8 & S_{l,r}^9 & S_{l,r}^{10}
\end{bmatrix}
\begin{bmatrix}
\gamma_2\\ \gamma_3\\ \gamma_4\\ \gamma_5\\ b_2\\ b_3\\ b_4\\ b_5\\ b_6\\ b_7
\end{bmatrix} =
\begin{bmatrix}
S_{1,1}^{11}\\ \vdots\\ S_{1,r}^{11}\\ \vdots\\ S_{m,n}^{11}\\ \vdots\\ S_{l,r}^{11}
\end{bmatrix}, \tag{3.22}$$

where

$$S^1_{m,n} := 4G_n^2\left[1-\cos\left(\frac{2\pi}{G_n}\sin\theta_m\right)-\cos\left(\frac{2\pi}{G_n}\cos\theta_m\right)+\cos\left(\frac{2\pi}{G_n}\sin\theta_m\right)\cos\left(\frac{2\pi}{G_n}\cos\theta_m\right)\right],$$

$$S^2_{m,n} := G_n^2\left\{3-2\left[\cos\left(\frac{2\pi}{G_n}\sin\theta_m\right)+\cos\left(\frac{2\pi}{G_n}\cos\theta_m\right)\right]+\frac{1}{2}\left[\cos\left(\frac{4\pi}{G_n}\sin\theta_m\right)+\cos\left(\frac{4\pi}{G_n}\cos\theta_m\right)\right]\right\},$$

$$\begin{aligned}S^3_{m,n} := \frac{1}{2}G_n^2\Bigg\{&8+2\cos\left(\frac{4\pi}{G_n}\sin\theta_m\right)\cos\left(\frac{4\pi}{G_n}\cos\theta_m\right)-\cos\left(\frac{4\pi}{G_n}\sin\theta_m\right)-\cos\left(\frac{4\pi}{G_n}\cos\theta_m\right)\\&-4\left[\cos\left(\frac{2\pi}{G_n}\sin\theta_m\right)+\cos\left(\frac{2\pi}{G_n}\cos\theta_m\right)\right]\Bigg\},\end{aligned}$$

$$\begin{aligned}S^4_{m,n} := \frac{1}{2}G_n^2\Bigg\{&8+\cos\left(\frac{2\pi}{G_n}\sin\theta_m\right)\cos\left(\frac{4\pi}{G_n}\cos\theta_m\right)+\cos\left(\frac{4\pi}{G_n}\sin\theta_m\right)\cos\left(\frac{2\pi}{G_n}\cos\theta_m\right)\\&-5\left[\cos\left(\frac{2\pi}{G_n}\sin\theta_m\right)+\cos\left(\frac{2\pi}{G_n}\cos\theta_m\right)\right]\Bigg\},\end{aligned}$$

$$S^5_{m,n} := 2\pi^2\left[\cos\left(\frac{2\pi}{G_n}\sin\theta_m\right)+\cos\left(\frac{2\pi}{G_n}\cos\theta_m\right)-2\right],$$

$$S^6_{m,n} := 2\pi^2\left[\cos\left(\frac{4\pi}{G_n}\sin\theta_m\right)+\cos\left(\frac{4\pi}{G_n}\cos\theta_m\right)-2\right],$$

$$S^7_{m,n} := 4\pi^2\left[\cos\left(\frac{2\pi}{G_n}\sin\theta_m\right)\cos\left(\frac{2\pi}{G_n}\cos\theta_m\right)-1\right],$$

$$S^8_{m,n} := 4\pi^2\left[\cos\left(\frac{4\pi}{G_n}\sin\theta_m\right)\cos\left(\frac{4\pi}{G_n}\cos\theta_m\right)-1\right],$$

$$S^9_{m,n} := 2\pi^2\left[\cos\left(\frac{2\pi}{G_n}(2\sin\theta_m-\cos\theta_m)\right)+\cos\left(\frac{2\pi}{G_n}(\sin\theta_m+2\cos\theta_m)\right)-2\right],$$

$$S^{10}_{m,n} := 2\pi^2\left[\cos\left(\frac{2\pi}{G_n}(\sin\theta_m-2\cos\theta_m)\right)+\cos\left(\frac{2\pi}{G_n}(2\sin\theta_m+\cos\theta_m)\right)-2\right],$$

$$S^{11}_{m,n} := -4\pi^2+2G_n^2\left[2-\cos\left(\frac{2\pi}{G_n}\sin\theta_m\right)-\cos\left(\frac{2\pi}{G_n}\cos\theta_m\right)\right].$$

The coefficient matrix has $l\times r$ rows and 10 columns, thus (3.22) is an overdetermined system, which can be solved by least-squares method. Moreover, we observe that the optimal parameters given in [22] are roughly chosen, as only one group of parameters were obtained and used to the computation for different frequencies, velocities and step sizes. This may yield much numerical dispersion for large wavenumbers and variable $k(x, y)$ (see examples in Section 4). To reduce the numerical dispersion and improve the accuracy of the difference scheme, we propose the following rule.

Rule 3.5 (*Refined Choice Strategy*)**.**
Step 1. *Estimate the interval* $I_G := [G_{\min}, G_{\max}]$.
Step 2. *Choose* γ_i $(i\in\mathbb{N}_5)$ *and* b_j $(j\in\mathbb{N}_7)$ *such that*

$$\begin{aligned}&(\gamma_1,\ldots,\gamma_5,b_1,\ldots,b_7)\\&\quad=\arg\min\left\{\|J(\gamma_1,\ldots,\gamma_5,b_1,\ldots,b_7;\cdot,\cdot)\|_{\infty,I_G\times I_\theta}:\sum_{j=1}^{5}\gamma_j=1,\ \sum_{j=1}^{7}b_j=1\right\}.\end{aligned}\tag{3.23}$$

In general, we can estimate I_G by using *a priori* information before choosing parameters. For example, if the frequency $f\in[f_{\min},f_{\max}]$ and the velocity $v\in[v_{\min},v_{\max}]$ then for a given step size h we have $G_{\min}:=\frac{v_{\min}}{hf_{\max}}$ and $G_{\max}:=\frac{v_{\max}}{hf_{\min}}$. In Tables 1 and 2, we present some groups of refined optimal parameters.

Fig. 2 shows the normalized phase and group velocity curves for the rotated 9-point difference scheme (the rotated 9p), the 25-point difference scheme (3.10) with parameters (3.12) (the global 25p) and the 25-point difference scheme (3.10) with refined optimal parameters in Tables 1 and 2 (the refined 25p), respectively. It is easy to find that both the global 25p and the refined 25p have much improvement over the rotated 9p. Additionally, the refined 25p has much less

Table 1
Refined optimal parameters.

I_G	[2, 2.5]	[2.5, 3]	[3, 4]	[4, 5]
γ_1	0.047999963330863	0.147293330179725	0.183180998407207	0.242239308507107
γ_2	−0.005925928525878	0.055379772312106	0.137353578227016	0.160596541280590
γ_3	0.413091635341205	0.373678865083601	0.360119886369911	0.327023199880231
γ_4	0.060680766447008	0.033859215177640	0.018241197835580	0.011791633533360
γ_5	0.484153563406802	0.389788817246928	0.301104339160286	0.258349316798712
b_1	0.220260676569402	0.282263140398365	0.309536538872974	0.341944671328177
b_2	0.432457082691776	0.441814591036386	0.472535719623776	0.469476235857888
b_3	0.042688088077091	0.026873864763897	0.022287994431260	0.017876897372456
b_4	0.249255472098946	0.216439606315222	0.174143429821568	0.154237668357416
b_5	0.003945393848657	0.001727236334600	−2.386237853810076e−4	−1.979300495582987e−4
b_6	0.025756984442826	0.015441838603192	0.010867906476688	0.008331249852036
b_7	0.025636302271301	0.015439722548337	0.010867034559120	0.008331207281584

Table 2
Refined optimal parameters (continue).

I_G	[5, 6]	[6, 8]	[8, 10]	[10, 400]
γ_1	0.271782609062930	0.284158049895963	0.299001074252186	0.297046004487648
γ_2	0.168528263975194	0.180982896912852	0.184118800813270	0.199464088268622
γ_3	0.310427677250485	0.304028496720736	0.295761877237356	0.297395880484200
γ_4	0.009649524814674	0.007447303978096	0.006627125174128	0.004284766350832
γ_5	0.239611924896718	0.223383252492354	0.214491122523060	0.201809260408696
b_1	0.357993213516404	0.365234692044752	0.373254845373659	0.374319397679971
b_2	0.465956594566756	0.470297541934252	0.468262706183388	0.474948027209072
b_3	0.016031996569628	0.015198470230312	0.014388012687780	0.015492621083616
b_4	0.145493965355276	0.136125397562404	0.131767373961080	0.125185968578920
b_5	−1.621668629146452e−4	−2.746678031531637e−4	−2.541135299981671e−4	−2.455072713816700e−4
b_6	0.007343200878844	0.006709277150948	0.006290587218416	0.005149658844840
b_7	0.007343195976008	0.006709288880488	0.006290588105672	0.005149833874960

numerical dispersion than the global 25p. Therefore, we expect that the refined 25p will be the best choice to control the numerical dispersion and to suppress non-physical oscillations. This will be illustrated by numerical experiments in the next section.

3.3. A generalization

In practice, we usually compute on the nonuniform grids. In this subsection, we generalize the consistent 25-point scheme to a nonuniform case, and show how to optimize the corresponding parameters in the interior domain.

To construct the finite difference scheme, we consider the network of grid points (x_m, y_n), where $x_m := x_0 + (m-1)\Delta x$ and $y_n := y_0 + (n-1)\Delta y$, with Δx, Δy being the step size of the variable x, y respectively. We denote

$$\tilde{\mathcal{L}}_{\Delta x,x} u\,|_{(m,n+j)} := \frac{A_{m+\frac{1}{2},n+j}(u_{m+1,n+j} - u_{m,n+j}) - A_{m-\frac{1}{2},n+j}(u_{m,n+j} - u_{m-1,n+j})}{(\Delta x)^2}, \tag{3.24}$$

for $j \in \mathbb{Z}_2 := \{-1, 0, 1\}$. It is easy to find that simply replacing the step size h with Δx in the left hand side of the Eq. (3.1) yields the left hand side of the above equation. Similarly, we introduce the notions $\tilde{\mathcal{L}}_{2\Delta x,x} u\,|_{(m,n+j)}$ for $j \in \mathbb{Z}_3$. Then we approximate the first term of the left hand side of (1.2) as

$$\frac{\partial}{\partial x}\left(A\frac{\partial u}{\partial x}\right)|_{x=x_m,y=y_n} \approx \tilde{\mathcal{L}}_{\Delta x,x} u\,|_{x=x_m,y=y_n} + \tilde{\mathcal{L}}_{2\Delta x,x} u\,|_{x=x_m,y=y_n},$$

where notions $\tilde{\mathcal{L}}_{\Delta x,x} u\,|_{x=x_m,y=y_n}$, $\tilde{\mathcal{L}}_{2\Delta x,x} u\,|_{x=x_m,y=y_n}$ are introduced following the way of definitions of $\tilde{\mathcal{L}}_{h,x} u\,|_{x=x_m,y=y_n}$, $\tilde{\mathcal{L}}_{2h,x} u\,|_{x=x_m,y=y_n}$ respectively (see, Section 3.1). The approximation of the second term is dealt with similarly, that is,

$$\frac{\partial}{\partial y}\left(B\frac{\partial u}{\partial y}\right)|_{x=x_m,y=y_n} \approx \tilde{\mathcal{L}}_{\Delta y,y} u\,|_{x=x_m,y=y_n} + \tilde{\mathcal{L}}_{2\Delta y,y} u\,|_{x=x_m,y=y_n}.$$

Let $\tilde{\mathcal{L}}_{\Delta x,\Delta y} := \tilde{\mathcal{L}}_{\Delta x,x} + \tilde{\mathcal{L}}_{\Delta y,y} + \tilde{\mathcal{L}}_{2\Delta x,x} + \tilde{\mathcal{L}}_{2\Delta y,y}$. We obtain the following 25-point finite difference approximation for the Helmholtz-PML equation (1.2)

$$\tilde{\mathcal{L}}_{\Delta x,\Delta y} u\,|_{x=x_m,y=y_n} + \mathcal{I}_h(Ck^2 u)\,|_{x=x_m,y=y_n} = f_{m,n}. \tag{3.25}$$

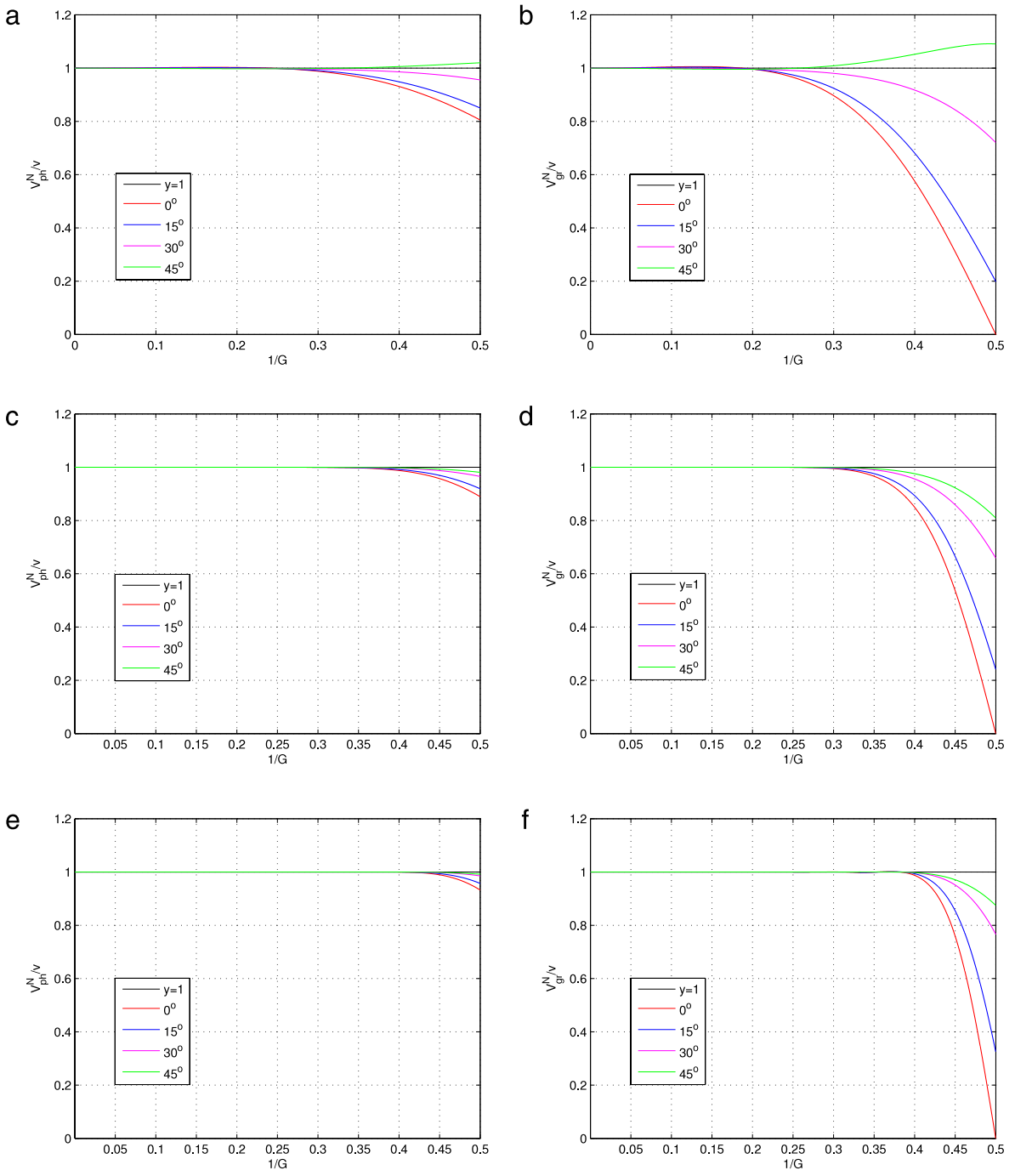


Fig. 2. (a) Normalized phase velocity curve for the optimal 9p, (b) Normalized group velocity curve for the optimal 9p, (c) Normalized phase velocity curve for the global 25p, (d) Normalized group velocity curve for the global 25p, (e) Normalized phase velocity curve for the refined 25p, (f) Normalized group velocity curve for the refined 25p.

Remark 3.6. If $\sum_{j=1}^{5}\gamma_j = 1$ and $\sum_{j=1}^{7} b_j = 1$, then the 25-point finite difference approximation (3.25) is consistent with the Helmholtz-PML equation (1.2).

When $\Delta x = h$, $\Delta y = \gamma h$ (γ is a positive constant), performing the classical dispersion analysis to the finite difference method (3.25), yields

$$k^N = \frac{1}{h}\sqrt{\frac{\tilde{R}}{\tilde{L}}}, \tag{3.26}$$

where

$$\tilde{L} := b_1 + \frac{b_2}{2}\left(P_1 + \tilde{Q}_1\right) + \frac{b_3}{2}\left(P_2 + \tilde{Q}_2\right) + b_4 P_1\tilde{Q}_1 + b_5 P_2\tilde{Q}_2 + \frac{b_6}{2}\left(\tilde{R}_1 + \tilde{R}_2\right) + \frac{b_7}{2}\left(\tilde{R}_3 + \tilde{R}_4\right),$$

$$\begin{aligned}\tilde{R} := {} & 2\gamma_1\left[\left(1+\frac{1}{\gamma^2}\right) - P_1 - \frac{1}{\gamma^2}\tilde{Q}_1\right] + 2\gamma_2\left[\frac{1}{\gamma^2}P_1 + \tilde{Q}_1 - \left(1+\frac{1}{\gamma^2}\right)P_1\tilde{Q}_1\right] + \frac{\gamma_3}{2}\left[\left(1+\frac{1}{\gamma^2}\right) - P_2 - \tilde{Q}_2\right] \\ & + \frac{\gamma_4}{2}\left[\frac{1}{\gamma^2}P_2 + \tilde{Q}_2 - \left(1+\frac{1}{\gamma^2}\right)P_2\tilde{Q}_2\right] + \frac{\gamma_5}{2}\left[\left(1-P_2\right)\tilde{Q}_1 + \frac{1}{\gamma^2}\left(1-\tilde{Q}_2\right)P_1\right],\end{aligned}$$

in which

$$\tilde{Q}_1 := \cos\left(\gamma kh\cos\theta\right) = \cos\left(\frac{2\pi}{G}\gamma\cos\theta\right), \qquad \tilde{Q}_2 := \cos\left(2\gamma kh\cos\theta\right) = \cos\left(\frac{4\pi}{G}\gamma\cos\theta\right),$$

$$\tilde{R}_1 := \cos\left(kh\left(2\sin\theta - \gamma\cos\theta\right)\right) = \cos\left(\frac{2\pi}{G}\left(2\sin\theta - \gamma\cos\theta\right)\right),$$

$$\tilde{R}_2 := \cos\left(kh\left(\sin\theta + 2\gamma\cos\theta\right)\right) = \cos\left(\frac{2\pi}{G}\left(\sin\theta + 2\gamma\cos\theta\right)\right),$$

$$\tilde{R}_3 := \cos\left(kh\left(\sin\theta - 2\gamma\cos\theta\right)\right) = \cos\left(\frac{2\pi}{G}\left(\sin\theta - 2\gamma\cos\theta\right)\right),$$

$$\tilde{R}_4 := \cos\left(kh\left(2\sin\theta + \gamma\cos\theta\right)\right) = \cos\left(\frac{2\pi}{G}\left(2\sin\theta + \gamma\cos\theta\right)\right).$$

As $h = \frac{2\pi}{Gk}$, we conclude that

$$\frac{k^N}{k} = \frac{G}{2\pi}\sqrt{\frac{\tilde{R}}{\tilde{L}}}. \tag{3.27}$$

Similarly, we choose optimal parameters γ_i ($i \in \mathbb{N}_5$) and b_j ($j \in \mathbb{N}_7$) by minimizing the numerical dispersion. To do this, we define the functional

$$J_\gamma(\gamma_1, \ldots, \gamma_5, b_1, \ldots, b_7; G, \theta) := \frac{G}{2\pi}\sqrt{\frac{\tilde{R}}{\tilde{L}}} - 1. \tag{3.28}$$

It follows from (3.27) that minimizing the error between the numerical wavenumber k^N and the exact wavenumber k is equivalent to minimizing the norm $\|J_\gamma(\gamma_1, \ldots, \gamma_5, b_1, \ldots, b_7; \cdot, \cdot)\|_{\infty, I_G, I_\theta}$. The corresponding refined choice strategy can be obtained by replacing $\|J(\gamma_1, \ldots, \gamma_5, b_1, \ldots, b_7; \cdot, \cdot)\|_{\infty, I_G \times I_\theta}$ in Rule 3.5 with $\|J_\gamma(\gamma_1, \ldots, \gamma_5, b_1, \ldots, b_7; \cdot, \cdot)\|_{\infty, I_G \times I_\theta}$.

Rule 3.7 (*Refined Choice Strategy II*)**.**

Step 1. *Estimate the interval* $I_G := [G_{\min}, G_{\max}]$.
Step 2. *Choose* γ_i ($i \in \mathbb{N}_5$) *and* b_j ($j \in \mathbb{N}_7$) *such that*

$$\begin{aligned}&(\gamma_1, \ldots, \gamma_5, b_1, \ldots, b_7) \\ &\quad = \arg\min\left\{\|J_\gamma(\gamma_1, \ldots, \gamma_5, b_1, \ldots, b_7; \cdot, \cdot)\|_{\infty, I_G \times I_\theta} : \sum_{j=1}^{5}\gamma_j = 1, \quad \sum_{j=1}^{7} b_j = 1\right\}.\end{aligned} \tag{3.29}$$

To implement Rule 3.7, we set $J_\gamma(\gamma_1, \ldots, \gamma_5, b_1, \ldots, b_7; G, \theta) = 0$ to obtain the equation

$$\begin{aligned}& G^2\left\{2\left(1+\frac{1}{\gamma^2}\right)\left(P_1\tilde{Q}_1 - \tilde{Q}_1 - P_1 + 1\right)\gamma_2 + \frac{1}{2}\left[3\left(1+\frac{1}{\gamma^2}\right) - 4P_1 - \frac{4}{\gamma^2}\tilde{Q}_1 + P_2 + \frac{1}{\gamma^2}\tilde{Q}_2\right]\gamma_3\right. \\ & \quad + \frac{1}{2}\left[\left(1+\frac{1}{\gamma^2}\right)P_2\tilde{Q}_2 - \frac{4}{\gamma^2}\tilde{Q}_1 - 4P_1 - \frac{1}{\gamma^2}P_2 - \tilde{Q}_2 + 4\left(1+\frac{1}{\gamma^2}\right)\right]\gamma_4 \\ & \quad \left. + \frac{1}{2}\left[\frac{1}{\gamma^2}P_1\tilde{Q}_2 + P_2\tilde{Q}_1 - \left(\frac{4}{\gamma^2}+1\right)\tilde{Q}_1 - \left(4+\frac{1}{\gamma^2}\right)P_1 + 4\left(1+\frac{1}{\gamma^2}\right)\right]\gamma_5\right\} \\ & \quad + 4\pi^2\left\{\frac{1}{2}\left(P_1 + \tilde{Q}_1 - 2\right)b_2 + \frac{1}{2}\left(P_2 + \tilde{Q}_2 - 2\right)b_3 + \left(P_1\tilde{Q}_1 - 1\right)b_4\right.\end{aligned}$$

Table 3
The error in the C-norm for $k = 200$.

N	129	257	513	1025
Rotated 9p	1.0258	0.4747	0.1298	0.0331
Global 25p	0.4326	0.0785	0.0210	0.0108
Refined 25p	0.4331	0.0788	0.0202	0.0052

$$+\left(P_2\tilde{Q}_2-1\right)b_5+\frac{1}{2}\left(\tilde{R}_1+\tilde{R}_2-2\right)b_6+\frac{1}{2}\left(\tilde{R}_3+\tilde{R}_4-2\right)b_7\Bigg\}$$
$$=-4\pi^2+2G^2\left\{\left(1+\frac{1}{\gamma^2}\right)-P_1-\frac{1}{\gamma^2}\tilde{Q}_1\right\}. \tag{3.30}$$

Solving Eq. (3.30) by the least square method as we did with Eq. (3.21) yields the refined optimal parameters for the finite difference scheme (3.25).

4. Numerical experiments

In this section, we give three numerical experiments to test our method proposed in the last section. The first one is an example for the Helmholtz equation, we show the advantage of the refined 25p over the rotated 9p and the global 25p by comparing the accuracy of these three schemes. The second one is a homogeneous model. We investigate the refined 25p's ability to portray the wave with a large step h by making comparisons among the rotated 9p, the global 25p and the refined 25p. Finally, a concave model is given to test the refined 25p for heterogeneous medium.

4.1. A numerical example for the Helmholtz equation

Consider the Helmholtz equation

$$-\Delta u-k^2u=0,\quad \text{in } \Omega:=(0,1)\times(0,1), \tag{4.1}$$

with boundary conditions

$$iku+\frac{\partial u}{\partial n}=g,\quad \text{on } \Gamma:=\partial\Omega. \tag{4.2}$$

The function g depends on the parameter θ, and is given by

$$g(x):=\begin{cases}\mathrm{i}(k-k_2)\exp(\mathrm{i}k_1x_1), & \text{if } x\in\Gamma_1:=(0,1)\times(0,0),\\ \mathrm{i}(k+k_1)\exp(\mathrm{i}(k_1+k_2x_2)), & \text{if } x\in\Gamma_2:=(1,1)\times(0,1),\\ \mathrm{i}(k+k_2)\exp(\mathrm{i}(k_1x_1+k_2)), & \text{if } x\in\Gamma_3:=(1,0)\times(1,1),\\ \mathrm{i}(k-k_1)\exp(\mathrm{i}k_2x_2), & \text{if } x\in\Gamma_4:=(0,0)\times(1,0),\end{cases}$$

with $(k_1,k_2):=k(\cos\theta,\sin\theta)$. The exact solution of this problem is

$$u(x):=\exp(\mathrm{i}(k_1x_1+k_2x_2)).$$

This problem was used for measuring the efficiency of numerical methods in [25]. We use it to compare the accuracy of the rotated 9p, the global 25p and the refined 25p. The error is measured in C-norm, which is defined as follows. For any complex vector $\mathbf{z}=[z_1,z_2,\ldots,z_M]$,

$$\|\mathbf{z}\|_C:=\max_{1\le j\le M}|z_j|,$$

where $|z_j|$ is the complex modulus of z_j. The parameters in Tables 1 and 2 are used as refined optimal parameters. Additionally, the 25-point finite difference schemes use the fourth order interpolation polynomials for the numerical boundary conditions (cf. [34]).

In Table 3, we present errors of these three schemes in C-norm for $k=200$ and $\theta=\frac{\pi}{4}$ with different gridpoints N per line. We find that, to obtain a certain accuracy, the refined 25p and the global 25p only need less than half of the grids that the rotated 9p needs, which is consistent with the dispersion analysis in Section 3.2. Also, we see that both the rotated 9p and the refined 25p are second order, while the global 25p seems a little slower than the second order, which may be caused by the pollution effect.

To further compare the three schemes in terms of accuracy, in Table 4, we show the error in the C-norm for the case of $\theta=\frac{\pi}{4}$ and $N=1025$ with $k=100,200,300,400$. From this table, we see that there is much improvement in accuracy for the global 25p and the refined 25p over the rotated 9p, especially for large wavenumbers. Additionally, there is an improvement for the refined 25p over the global 25p, though it is not so sharp.

All of the experiments above are done with $\theta=\frac{\pi}{4}$. We finally test these schemes' dependence on θ. Table 5 shows the corresponding results for $k=400$ and $N=1025$, with θ varying among $[0,\pi/4]$. It is clear that, the refined 25p and the

Table 4
The error in the C-norm for $N = 1025$.

k	100	200	300	400
Rotated 9p	0.0043	0.0331	0.1108	0.2599
Global 25p	0.0049	0.0108	0.0175	0.0251
Refined 25p	0.0013	0.0052	0.0115	0.0202

Table 5
The error in the C-norm for $k = 400$ and $N = 1025$.

θ	0	$\frac{\pi}{16}$	$\frac{\pi}{8}$	$\frac{3\pi}{16}$	$\frac{\pi}{4}$
Rotated 9p	0.3519	0.4017	0.1419	0.1260	0.2599
Global 25p	0.0313	0.0407	0.0326	0.0275	0.0251
Refined 25p	0.0313	0.0403	0.0313	0.0247	0.0202

global 25p's dependence on the wave direction θ are much less than the rotated 9p and their accuracy is much higher than that of the rotated 9p. The refined 25p also has an improvement in accuracy over the global 25p. In practice, we usually do not know the wave's exact propagation direction, so we had better choose the refined 25p.

From this example, we know that the refined 25p exactly possess advantages over both the rotated 9p and the global 25p, which demonstrates that our proposed refined 25p is useful and promising.

4.2. Constant wavenumber

To test the optimal 25-point schemes for the Helmholtz-PML equation (1.2), we use a 2-D homogeneous model, as shown in Fig. 4(a). The velocity of the medium is 1500 m/s. A point source $\delta(x - x_s, y - y_s)R(\omega, f_0)$ is located at the point $(x_s, y_s) = (500\text{ m}, 1000\text{ m})$, where $R(\omega, f_0)$ is the Ricker wavelet with

$$R(t, f_0) = (1 - 2\pi^2 f_0^2 t^2)\exp(-\pi^2 f_0^2 t^2),$$

whose dominant frequency is $f_0 = 25$ Hz. The grid size is 10 m, the time sampling is 8 ms and the highest frequency we compute is 60 Hz.

For 201×201 mesh points, we generate synthetic seismograms by the extended 25p, the optimal 9-point difference scheme (the optimal 9p) proposed in [13], the global 25p and the refined 25p. Figs. 4(b) and 3 show the synthetic seismograms for each formulation. Seen from Fig. 4(b), we find many non-physical oscillations appearing in the synthetic seismogram obtained by the extended 25p. From Fig. 3(b)–(d), we see that the synthetic seismogram based on the optimal 9p suffers from much more non-physical oscillations than the global 25p, while there are almost no non-physical oscillations in the synthetic seismogram generated by the refined 25p. In addition, the maximum value in Fig. 3(b)–(d) appears at the same time of 0.6720 s. According to the velocity of the wave, the wave propagates a distance of 1008 m, which is an appropriate result in practice. Moreover, snapshots at the time $t = 256$ ms generated by the optimal 9p and the refined 25p are shown in Fig. 3(e) and (f) respectively. As the model is a homogeneous model, the exact snapshot is a perfect circle without any reflections when the wave does not arrive at the boundaries. However, we see that the snapshot generated by the optimal 9p suffers from many false reflections, while the snapshot based on the refined 25p is very clear.

Now, we come to our conclusion that the refined 25p is the best choice as it has simulated the wave propagation better and suppresses false reflections much more effectively when compared with the optimal 9p and the global 25p.

4.3. Concave model

The concave model is shown in Fig. 5(a). There are three velocities in this model: from the top, $v = 1500$ m/s, 2000 m/s, 2500 m/s. The grid size is taken as $\Delta x = \Delta y := 10$ m and the interval of time sampling $\Delta t := 8$ ms. The source is located at the point (1000 m, 800 m). As in the homogeneous model above, the Ricker wavelet with domain frequency 25 Hz is used as the source.

For 201×201 mesh points, we generate snapshots for the wavefield by the optimal 9p and the refined 25p. The monofrequency wavefield (real part) for $f = 25$ Hz obtained by the refined 25p is present in Fig. 5(b). The PML's absorption is efficient, and almost no boundary reflections exist. Additionally, the upward incident waves, the downward incident waves and transmissive waves are all clear. We also note that in the middle velocity layer, the incident waves are interfered with the reflected waves returned from the middle velocity layer's interfaces, and this phenomena obeys Snell's law. In Fig. 5(c), we show the snapshot for the time being 400 ms, generated by the optimal 9-point scheme, and Fig. 5(d) shows the corresponding snapshot generated by the refined 25p. Comparing Fig. 5(d) with Fig. 5(c), we find that Fig. 5(d) is clearer than Fig. 5(c), indicating that the refined 25p has less numerical dispersion than the optimal 9p.

5. Conclusion

In this section, we summarize and comment on the difference methods. In Section 2, we proved that, though it is a popular scheme for the Helmholtz equation, the extended 25-point finite difference scheme with its optimal parameters

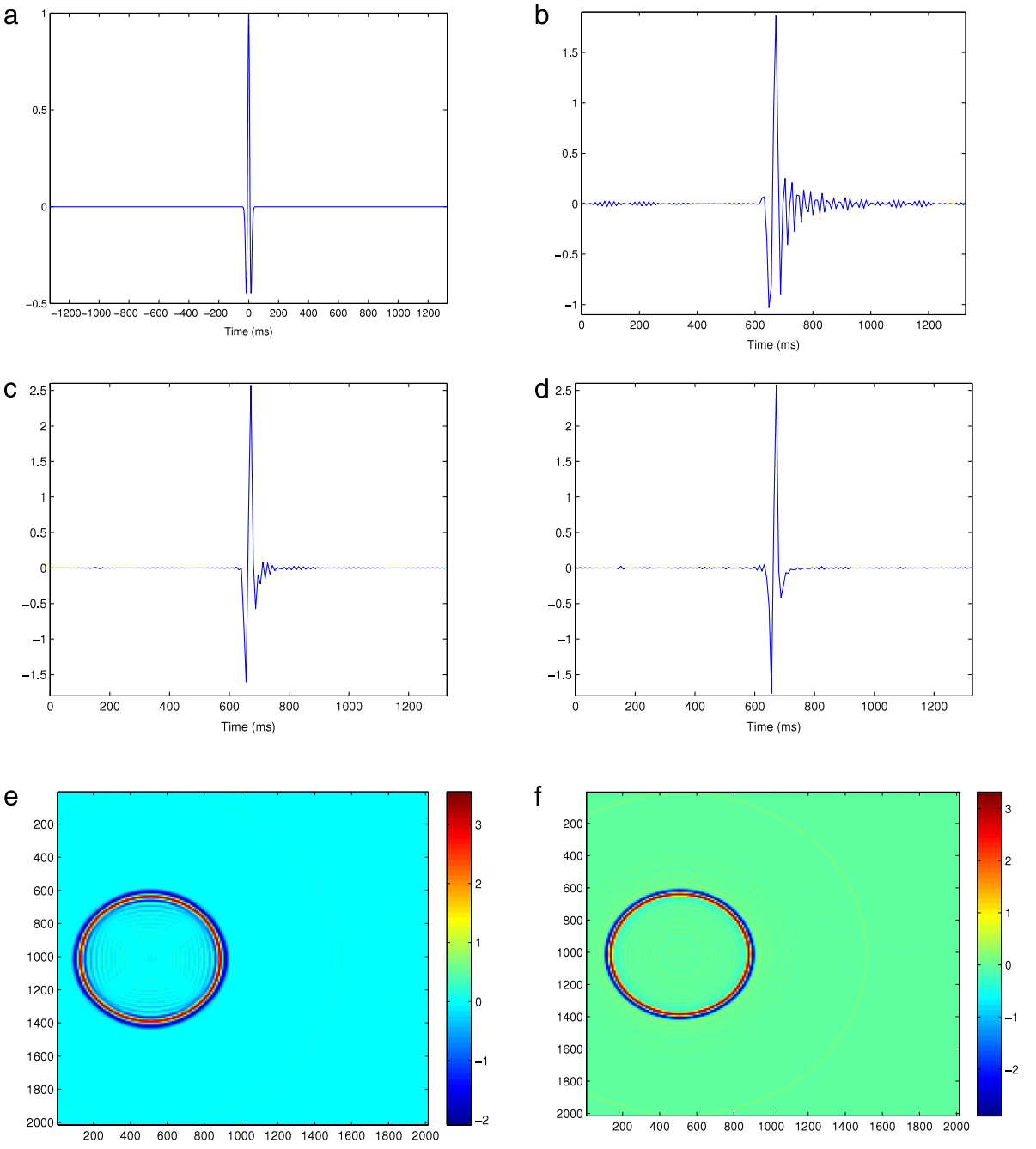


Fig. 3. (a) Ricker wavelet, (b) Synthetic seismograms generated by the optimal 9p, (c) Synthetic seismograms generated by the global 25p, (d) Synthetic seismograms generated by the refined 25p, (e) Snapshot generated by the optimal 9p at the time $t = 256$ ms, (f) Snapshot generated by the refined 25p at the time $t = 256$ ms.

given in [22] is not pointwise consistent with the Helmholtz-PML equation. To deal with this problem, a consistent 25-point difference scheme was constructed and the refined choice strategy for choosing optimal parameters based on minimizing the numerical dispersion were proposed in Section 3. Additionally, normalized numerical phase and group velocity curves showed that the refined 25-point finite difference scheme has the smallest numerical dispersion and anisotropy, compared with the extended 25-point finite difference scheme and the optimal 9-point finite difference scheme. Finally, numerical experiments in Section 4 were presented to confirm that the refined 25-point finite difference scheme is the best scheme for the Helmholtz-PML equation. The applications ranged from constant wavenumbers to varying wavenumbers. Compared to the optimal 9-point finite difference scheme and the global 25-point finite difference scheme, the refined 25-point

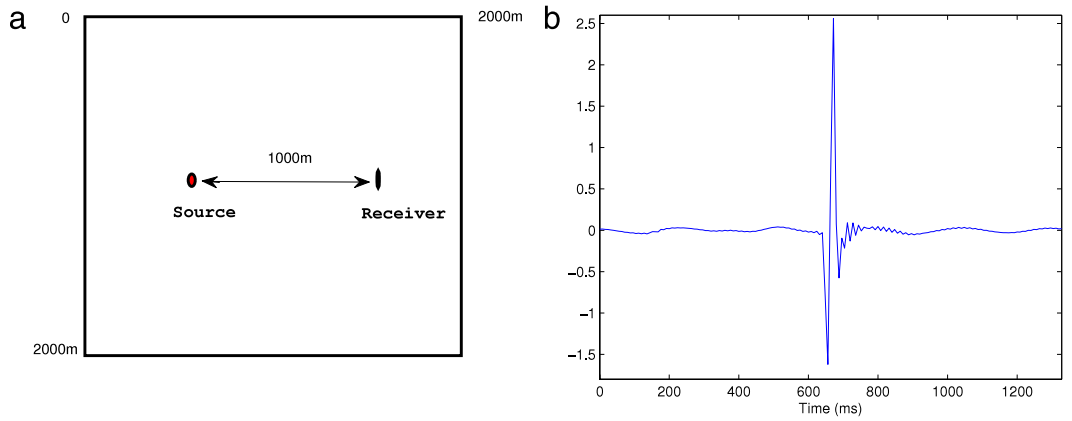


Fig. 4. (a) The homogeneous model, (b) Synthetic seismograms generated by the extended 25p.

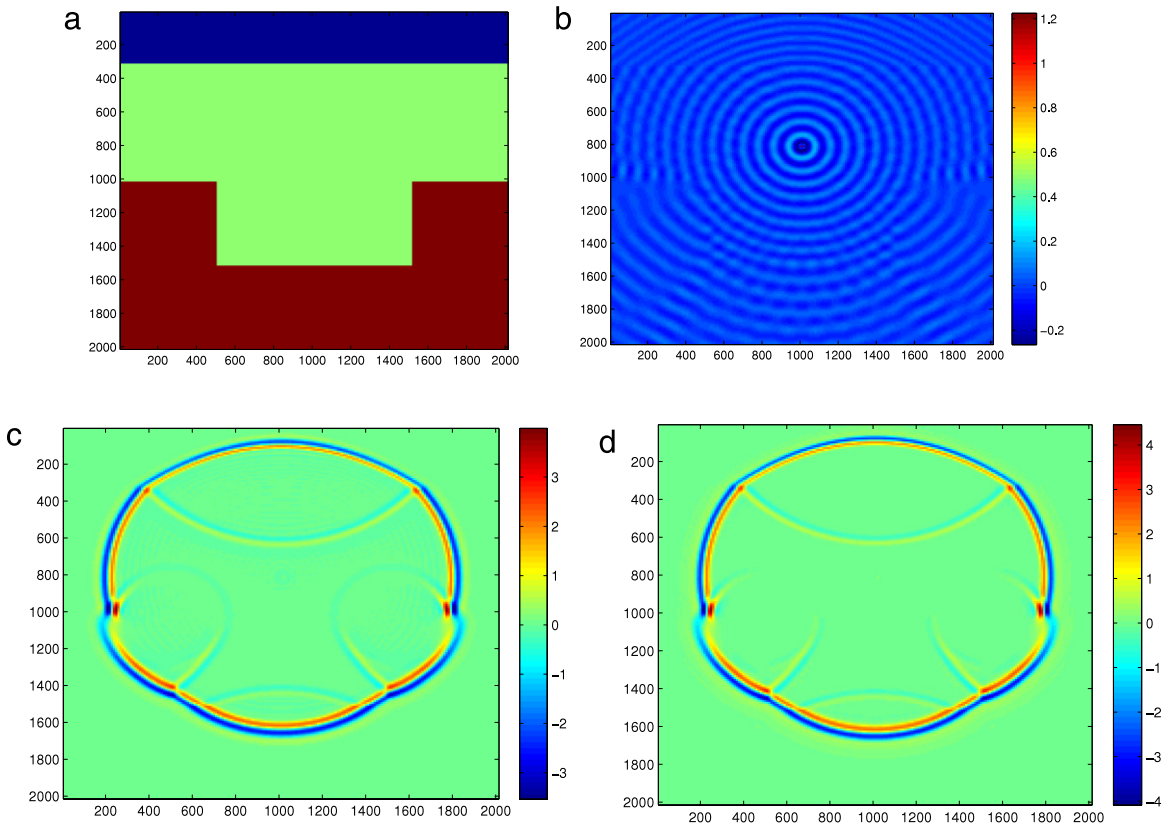


Fig. 5. (a) Concave model, (b) Monofrequency wavefield (real part) for $f = 25$ Hz obtained by the refined 25p, (c) Snapshot generated by the optimal 9p at the time $t = 400$ ms, (d) Snapshot generated by the refined 25p at the time $t = 400$ ms.

finite difference scheme possesses the ability of portraying waves well with large steps while suppressing non-physical oscillations most effectively, especially for large wavenumbers.

References

[1] R.G. Pratt, C. Shin, G.J. Hicks, Gauss–Newton and full Newton methods in frequency-space seismic waveform inversion, Geophys. J. Int. 133 (1998) 341–362.
[2] R.G. Pratt, M.H. Worthington, Inverse theory applied to multi-source cross-hole tompgraphy. Part 1: acoustic wave-equation method, Geophys. Prospect. 38 (1990) 287–310.
[3] R. Clayton, B. Engquist, Absorbing boundary conditions for acoustic and elastic wave equations, Bull. Seismol. Soc. Amer. 67 (1977) 1529–1540.
[4] R.W. Clayton, B. Engquist, Absorbing boundary conditions for wave-equations migration, Geophysics 45 (1980) 895–904.
[5] B. Engquist, A. Majda, Absorbing boundary conditions for the numerical simulation of waves, Math. Comp. 31 (1977) 629–651.

[6] J.-P. Bérenger, A perfectly matched layer for the absorption of electromagnetic waves, J. Comput. Phys. 114 (1994) 185–200.
[7] W.C. Chew, J.M. Jin, E. Michielssen, Complex coordinate stretching as a generalized absorbing boundary condition, Microw. Opt. Technol. Lett. 15 (1997) 363–369.
[8] F.D. Hastings, J.B. Schneider, S.L. Broschat, Application of the perfectly matched layer (PML) absorbing boundary condition to elastic wave propagation, J. Acoust. Soc. Am. 100 (1996) 3061–3069.
[9] I. Singer, E. Turkel, A perfectly matched layer for the Helmholtz equation in a semi-infinite strip, J. Comput. Phys. 201 (2004) 439–465.
[10] S. Tsynkov, E. Turkel, A Cartesian perfectly matched layer for the Helmholtz equation, in: Absorbing Boundaries and Layers, Domain Decomposition Methods, Nova Sci. Publ., Huntington, NY, 2001, pp. 279–309.
[11] E. Turkel, A. Yefet, Absorbing PML boundary layers for wave-like equations, Appl. Numer. Math. 27 (1998) 533–557.
[12] Y. Zeng, J. He, Q. Liu, The application of the perfectly matched layer in numerical modeling of wave propagation in poroelastic media, Geophysics 66 (2001) 1258–1266.
[13] Z. Chen, D. Cheng, W. Feng, T. Wu, An optimal 9-point finite difference scheme for the Helmholtz equation with PML, Preprint.
[14] J.W. Kang, L.F. Kallivokas, Mixed unsplit-field perfectly matched layers for transient simulations of scalar waves in heterogeneous domains, Comput. Geosci. 14 (2010) 623–648.
[15] J.W. Kang, L.F. Kallivokas, The inverse medium problem in heterogeneous PML-truncated domains using scalar probing waves, Comput. Methods Appl. Mech. Engrg., 200 (2011) 265–283.
[16] G. Baruch, G. Fibich, S. Tsynkov, E. Turkel, Fourth order schemes for time-harmonic wave equations with discontinuous coefficients, Commun. Comput. Phys. 5 (2009) 442–455.
[17] Y.A. Erlangga, C.W. Oosterlee, C. Vuik, A novel multigrid based preconditioner for heterogeneous Helmholtz problems, SIAM J. Sci. Comput. 27 (2006) 1471–1492.
[18] I. Harari, E. Turkel, Accurate finite difference methods for time-harmonic wave propagation, J. Comput. Phys. 119 (1995) 252–270.
[19] B. Hustedt, S. Operto, J. Virieux, Mixed-grid and staggered-grid finite-difference methods for frequency-domain acoustic wave modelling, Geophys. J. Int. 157 (2004) 1269–1296.
[20] C.-H. Jo, C. Shin, J.H. Suh, An optimal 9-point, finite-difference, frequency-space, 2-D scalar wave extrapolator, Geophysics 61 (1996) 529–537.
[21] H. Ren, H. Wang, T. Gong, Seismic modeling of scalar seismic wave propagation with finite difference scheme in frequency-space domain, Geophys. Prospect. Pet. 48 (2009) 20–27.
[22] C. Shin, H. Sohn, A frequency-space 2-D scalar wave extrapolator using extended 25-point finite-difference operator, Geophysics 63 (1998) 289–296.
[23] I. Singer, E. Turkel, High-order finite difference methods for the Helmholtz equation, Comput. Methods Appl. Mech. Engrg. 163 (1998) 343–358.
[24] M.B. Van Gijzen, Y.A. Erlangga, C. Vuik, Spectral analysis of the discrete Helmholtz operator preconditioned with a shifted laplacian, SIAM J. Sci. Comput. 29 (2007) 1942–1958.
[25] I. Babuška, F. Ihlenburg, E.T. Paik, S.A. Sauter, A genralized finite element method for solving the Helmholtz equation in two dimensions with minimal pollution, Comput. Methods Appl. Mech. Engrg. 128 (1995) 325–359.
[26] I. Babuška, S.A. Sauter, Is the pollution effect of the FEM avoidable for the Helmholtz equation considering high wave numbers? SIAM Rev. 42 (2000) 451–484.
[27] A. Deraemaeker, I. Babuška, P. Bouillard, Dispersion and pollution of the FEM sollution for the Helmholtz equation in one, two and three dimensions, Internat. J. Numer. Methods Engrg. 46 (1999) 471–499.
[28] X. Feng, H. Wu, Discontinuous Galerkin methods for the Helmholtz equation with large wave number, SIAM J. Numer. Anal. 47 (2009) 2872–2896.
[29] F. Ihlenburg, I. Babuška, Finite element solution of the Helmholtz equation with high wave number, Part I: the h-version of the FEM, Comput. Math. Appl. 30 (1995) 9–37.
[30] Y. Guo, F. Ma, Some domain decomposition methods employing the PML technique for the Helmholtz equation, Numer. Math. J. Chinese Univ. 31 (2009) 369–384.
[31] P.M. De Zeeuw, Matrix-dependent prolongations and restrictions in a blackbox multigrid solver, J. Comput. Appl. Math. 33 (1990) 1–27.
[32] F. Ihlenburg, I. Babuška, Dispersion analysis and error estimation of galerkin finite element methods for the Helmholtz equation, Internat. J. Numer. Methods Engrg. 38 (1995) 4207–4235.
[33] Z. Chen, D. Cheng, W. Feng, T. Wu, H. Yang, A multigrid-based preconditioned Krylov subspace method for the Helmholtz equation with PML, J. Math. Anal. Appl. 383 (2011) 522–540.
[34] J.W. Thomas, Numerical Partial Differential Equations, Finite Difference Methods, Springer, New York, 1995.
[35] L.N. Trefethen, Group velocity in finite difference schemes, SIAM Rev. 24 (1982) 113–136.