ON OPTIMAL FINITE-DIFFERENCE APPROXIMATION OF PML

SERGEY ASVADUROV†, VLADIMIR DRUSKIN†, MURTHY N. GUDDATI‡, AND
LEONID KNIZHNERMAN§

Abstract. A technique derived from two related methods suggested earlier by some of the
authors for optimization of finite-difference grids and absorbing boundary conditions is applied to
discretization of perfectly matched layer (PML) absorbing boundary conditions for wave equations
in Cartesian coordinates. We formulate simple sufficient conditions for optimality and implement
them. It is found that the minimal error can be achieved using pure imaginary coordinate stretching.
As such, the PML discretization is algebraically equivalent to the rational approximation of the
square root on [0,1] conventionally used for approximate absorbing boundary conditions. We present
optimal solutions for two cost functions, with exponential (and exponential of the square root) rates
of convergence with respect to the number of the discrete PML layers using a second order finite-
difference scheme with optimal grids. Results of numerical calculations are presented.

Key words. absorbing boundary conditions, exponential convergence, finite differences, hyper-
bolic problems, perfectly matched layers, wave propagation, optimal rational approximations

AMS subject classifications. 65N06, 73C02

1. Introduction. This paper is a sequel to a number of papers on so-called opti-
mal finite-difference grids or finite-difference Gaussian rules [11, 12, 3, 4, 18, 13], where
exponential superconvergence of standard second order finite-difference schemes at a
priori given points was obtained due to a special grid optimization procedure. This
approach was successfully applied to the approximation of many nontrivial practi-
cally important problems, including elliptic PDEs for both bounded and unbounded
domains. For the latter, in [18] the optimal finite-difference grid was obtained, which
can be considered as the boundary condition requiring minimal arithmetic work for a
given spectral interval. For hyperbolic problems, however, optimal grids were intro-
duced only for the approximation in the interior part of the domain. Here we consider
exterior hyperbolic problems. For this kind of problem a closely related method of con-
tinued fraction boundary conditions was suggested in [16], where absorbing boundary
conditions were reduced to a three-term equations resembling finite-difference rela-
tions. Combining the approaches of [18] and [16] we obtain frequency independent
finite-difference discretization of Berenger’s perfectly matched layer (PML) absorbing
boundary conditions (ABCs) [7], which produces the minimal possible impedance er-
ror for a given number of discrete layers. Similarly to the optimal grid for the Laplace
equation with a solution from a Sobolev space considered in [18], the obtained discret-
izations show the exponential of the square root rates of convergence, though they
use only the three-point stencil for second derivatives. Our solution exhibits much
smaller reflection coefficients compared to examples of optimized PMLs (for the same
numbers of discrete layers) known from the literature [10]. The optimal solution represents a limiting case of PML with pure imaginary coordinate stretching that results in a different type of the equation in time domain. The drawback of the proposed method is that, unlike a conventional PML discretization, it cannot be implemented in time domain using the same time-stepping realization in the interior and the exterior parts. Nevertheless, the arithmetical cost per grid node of the new method is close to that of the conventional PML.

Let us consider a model problem for the scalar wave equation on \( \mathbb{R}^2 \times [0, +\infty) \) in Cartesian coordinates

\[ u_{xx} + u_{yy} - u_{tt} = 0. \]  

(1.1)

After the Fourier transform with respect to \( y \) and \( t \), (1.1) becomes

\[ u_{xx} - (l^2 - \omega^2) u = 0, \]  

(1.2)

where \( l \) and \( \omega \) are real spatial and temporal frequencies, respectively. We here abuse notation slightly by using the same name for a function and its Fourier transform; the use of the time or frequency domain will always be clear from the context. The ratio \( \sigma = l/\omega \) is the sine of the incidence angle of the wave on the plane \( x = 0 \); this angle is labelled \( \theta \).

We assume that only so-called propagative modes with \( |\sigma| \leq 1 \) are present in the spectrum of the solution and that for positive \( x \) the solution contains only waves moving to the right; i.e., we are considering solutions of the form

\[ u(x) = ce^{-i\omega\sqrt{\lambda}x} \text{ if } x > 0, \]  

(1.3)

where \( \lambda = 1 - \sigma^2 = \cos^2 \theta \) is positive.

The solutions of this form do not vanish at infinity, which is the origin of the notorious problem of domain truncation in the numerical solution of wave problems on unbounded domains. Functions given by (1.3) satisfy the impedance boundary condition

\[ u_x\big|_{x=0} = -i\omega\sqrt{\lambda}u\big|_{x=0}. \]  

(1.4)

With the help of this condition, the subdomain \( x > 0 \) can be truncated. Many approximate ABCs are based on rational approximation of \( \sqrt{\lambda} \), e.g., in [14, 16]. The authors of these investigations used the fact that the approximant after the inverse Fourier transform to \( (y, t) \) coordinates becomes a solution of a PDE in \( (y, t) \) plane; i.e., it can be computed within the finite-difference time-stepping framework of the solution of (1.1).

Instead of the direct implementation of the approximate condition (1.4), one can modify (1.2) for \( x > 0 \) in such a way that it would be easier to solve and the new solution would approximate (1.4) well. The methods of this sort that recently received wide attention in the literature, e.g., in [7, 9, 10, 19], are Berenger’s PML or sponge layer methods. These methods generate special artificial media layers that add exponential decay or attenuation to the propagative modes so that the new solution satisfies the same boundary condition at \( x = 0 \). Thus, the domain for the new equation can be truncated, which will produce only a small reflection for \( x < 0 \). For brevity we will call PML all methods from this group.
It is known that Berenger’s PML can be obtained using complex coordinate stretching [9]:

\[ dx = \left( \alpha + \frac{\beta}{i\omega} \right) d\bar{x} \quad \text{if} \ x > 0, \quad x = \bar{x} \quad \text{otherwise}, \]

where \( \alpha \) and \( \beta \) are some real nonnegative functions. The \( \omega^{-1} \) dependence of the imaginary part is introduced to simplify the time-domain formulation and weaken the dependence on \( \omega \) of the PML error. The new function \( \bar{u}(\bar{x}) = u(x) \) is defined, and the equation is modified in the following way: the new variable \( \bar{x} \) is taken to be real. This transformation does not change the solution for negative \( \bar{x} \), and for positive \( \bar{x} \) it transforms (1.3) to

\[ \bar{u}(\bar{x}) = c \exp \left[ -\sqrt{\lambda} \int_{0}^{\bar{x}} \beta(\xi) d\xi \right] \exp \left[ -i\omega \sqrt{\lambda} \int_{0}^{\bar{x}} \alpha(\xi) d\xi \right]; \]

i.e., the exponential attenuation is added to the resulting function. Because of this attenuation of the solution in the new coordinate, the subdomain \([0, +\infty)\) can be truncated to a finite length \( L \), with the logarithm of the absolute value of the impedance error (or reflection coefficient) approximately proportional to

\[ R = -\sqrt{\lambda} \int_{0}^{L} \beta(\xi) d\xi. \]

From this estimate one might conclude that just choosing large enough \( \beta \) alone would make error negligibly small, regardless of the temporal frequency. Unfortunately, this consideration cannot be applied to discretized PML because of the numerical dispersion, which is frequency dependent and increases with the increase of \( \beta \). Conventionally, ad hoc rules or general nonlinear optimization algorithms have been used for the choice of the discrete PML parameters, e.g., in [10]. It is obvious that proper use of the analytical structure of the cost function can greatly improve the efficiency of such optimization; this is the main motivation of our investigation.

Let \( u^{k} \) denote the solution of (1.2) after a three-point second order finite-difference discretization of PML with \( k \) primary and \( k \) dual nodes. In this case, \( \alpha \) and \( \beta \) become finite \( 2k \)-dimensional vectors that determine the discrete transformation at the primary and dual grid nodes. We consider the following PML optimization problem:

- Find \( \alpha, \beta \) minimizing the error functional

\[ \delta_{k}^{\alpha,\beta} = \sup_{0 \leq \lambda \leq 1, \ 0 < \omega \leq \omega_{\text{max}}} s(\lambda) \left| \sqrt{\lambda} - \frac{u_{x}^{k}}{i\omega u^{k}} \right|_{x=0}, \]

where \( u_{x}^{k} \) is the finite-difference derivative of the discrete solution at the boundary. Here we assume that the temporal spectrum of the solution is uniformly bounded at \((0, \omega_{\text{max}}]\) by a positive constant, where \( \omega_{\text{max}} \) is a cutoff frequency (as we shall see, the value of \( \omega_{\text{max}} \) is unimportant for the construction of our optimal solution), and \( s(\lambda) \) is a nonnegative weight, chosen depending on the distribution of the incident waves. We will consider two cases:

\[ s(\lambda) = \begin{cases} \frac{1}{\sqrt{\lambda}} & \text{if} \ \lambda \in [\lambda_{\text{min}}, 1], \lambda_{\text{min}} > 0, \\ 0 & \text{otherwise}, \end{cases} \]

and

\[ s(\lambda) = \begin{cases} 1 & \lambda \in [0, 1]. \end{cases} \]
Here the first case corresponds to all possible waves with the incidence angles not exceeding \( \arccos(\sqrt{\lambda_{\text{min}}}) < \frac{\pi}{2} \); the second case assumes that the wave amplitudes are uniformly bounded for all incidence angles on \([0, \pi/2]\). The first case will be used when a priori information limiting the range of incidence angles is available; the second case will generally be used when no such information can be obtained from the geometry of the model.

Traditionally, only equidistant grids were used for the PML discretization, and the optimization was performed by adjusting the distribution of \( \alpha \) or \( \beta \). Such discretization can be equivalently presented as the one on a nonuniform grid with complex grid steps but with \( \alpha = 1, \beta = 0 \). So, we can apply the approach of [11, 18], which reduces the problem of the finite-difference grid optimization to rational approximation and allows us to make the following conclusions, which will be proven in the following sections:

- The minimum of \( \delta_{k}^{\alpha,\beta} \) can be achieved with pure imaginary stretching:
  \[
  \min_{\alpha,\beta} \delta_{k}^{\alpha,\beta} = \min_{\beta} \delta_{k}^{0,\beta},
  \]
  though it is not clear if the condition \( \alpha \equiv 0 \) is necessary for optimality.
- With pure imaginary stretching the PML discretization becomes independent of \( \omega \) and depends only on the incidence angle. In other words, the PML discretization becomes a well-studied problem of rational approximation of the \( \sqrt{\lambda} \) on a nonnegative interval of real axis.

Choosing appropriate grid steps (or \( \beta \)) for PML with pure imaginary stretching, one can make it algebraically equivalent to known approximate ABCs of, e.g., [14, 16]. Instead, using known results of approximation theory, we give optimal solutions for the two weight functions considered here, which are efficient for wide bandwidth of the incidence angles. For a problem similar to (1.7) the optimal PML discretization is based on a closed form solution, obtained in 1877 by Zolotarjov [20]. If \( 0 < \lambda_{\text{min}} \ll 1 \) its error decays with the increase of \( k \) approximately as

\[
(1.9)
O \left\{ \exp \left[ \frac{\pi^2 k}{\log \left( \sqrt{\lambda_{\text{min}}}/4 \right)} \right] \right\}.
\]

For the case (1.8) the optimal grid was computed in the manner of [24], where optimal rational approximants of a slightly different type were computed to very high precision. For the latter case the optimal error asymptotically decays as

\[
(1.10)
8 e^{-\pi \sqrt{2k}}.
\]

Both of these estimates again highlight the phenomenon of exponential superconvergence of three-point second order finite-different approximations, which has been earlier used for efficient approximation of elliptic and hyperbolic problems in [11, 3, 18, 12].

The condition \( \alpha \equiv 0 \) does not allow standard split time-domain PML realization; however, it makes possible a simple nonsplit realization, which is, in fact, similar to the one of [16].

2. S-fraction representation of discrete PML. We want the propagating solution \( u = ce^{-i \sqrt{\omega^2 - \lambda} x} \) to become nonoscillating evanescent for \( x > 0 \). We consider a limiting case of transformation (1.5) with \( \alpha = 0 \) and \( \beta = 1 \):

\[
\bar{x} = i \omega x \text{ if } x > 0, \quad \bar{x} = x \text{ otherwise}.
\]

\[
(1.11)\]
Define a new function \( \tilde{u}(\tilde{x}) = u(x) \) and again take \( \tilde{x} \) to be real. The equation that is satisfied by this new function can be written in divergence form as
\[
\frac{d}{d\tilde{x}} \left[ \gamma(\tilde{x}) \frac{d\tilde{u}(\tilde{x})}{d\tilde{x}} \right] - \lambda \rho(\tilde{x}) \tilde{u}(\tilde{x}) = 0,
\]
where \( \lambda = 1 - l^2/\omega^2 = \cos^2 \theta \) and
\[
\gamma = i\omega, \quad \rho = i\omega \text{ if } \tilde{x} > 0, \quad \gamma = 1, \quad \rho = -\omega^2 \text{ otherwise}.
\]
Equation (2.1) is a standard divergence equation with discontinuous coefficients; \( \tilde{u}(\tilde{x}) \) and \( \gamma(\tilde{x}) \frac{d\tilde{u}(\tilde{x})}{d\tilde{x}} \) are continuous across the interface at \( \tilde{x} = 0 \).

The transformed Helmholtz equation (2.1) becomes diffusive (absorbing) for \( \tilde{x} > 0 \) and remains the same as the original oscillating (1.2) otherwise. Its solution vanishing at \(+\infty\) can be written as
\[
\tilde{u} = ce^{-\sqrt{\lambda} \tilde{x}} \text{ if } \tilde{x} > 0, \quad \tilde{u} = ce^{-i\omega \sqrt{\lambda} \tilde{x}} = u \text{ otherwise}.
\]
From now on we will approximate the above solution for \( \tilde{x} \in \mathbb{R} \), so we will drop the bars over all the symbols. The impedance condition (1.4) is then written as
\[
|u|_{x=0+} = -\sqrt{\lambda} |u|_{x=0+},
\]
and the error functional (1.6) for the discrete absorbing condition is
\[
\delta_k = \max_{0 \leq \lambda \leq 1} s(\lambda) \left| \sqrt{\lambda} - \frac{u_k}{u_k}|_{x=0} \right|,
\]
where \( u_k \) is the discrete solution of (2.1), and \( u_k \) stands for the finite-difference analogue of the derivative of \( u \).

Now, using a second order finite-difference approximation of the diffusive part of (2.1) as it was done for diffusive problems in [11, 18], we obtain a formulation equivalent to the continued fraction of [16].

We want to approximate the ABC (2.3) and to get an explicit expression for the error functional (2.4) in terms of the steps of the finite-difference discretization. For this purpose we first consider (2.1) for \( x \in [0, +\infty) \) with the following boundary conditions:
\[
-\gamma u'(0) = 1, \quad u(+\infty) = 0.
\]
Let us approximate the solution of this problem by a staggered three-point finite-difference scheme. In a staggered scheme, the numerical solution is defined at primary nodes
\[
x_j, \quad j = 1, \ldots, k + 1, \quad \text{with} \quad x_1 = 0 \quad \text{and} \quad x_{j+1} > x_j \ (1 \leq j \leq k),
\]
and the finite-difference derivatives are defined at dual nodes
\[
\hat{x}_j, \quad j = 0, \ldots, k, \quad \text{with} \quad \hat{x}_0 = 0 \quad \text{and} \quad \hat{x}_{j+1} > \hat{x}_j \ (1 \leq j \leq k - 1).
\]
We denote the step sizes by
\[
h_j = x_{j+1} - x_j \quad \text{and} \quad \hat{h}_j = \hat{x}_j - \hat{x}_{j-1}.
\]
and solve the following finite-difference problem:

\[
\frac{1}{h_j} \left( \gamma_j \frac{u_{j+1} - u_j}{h_j} - \gamma_{j-1} \frac{u_j - u_{j-1}}{h_{j-1}} \right) - \lambda \rho_j u_j = 0, \quad j = 2, \ldots, k,
\]

with the boundary conditions

\[
\frac{\hat{\gamma}_1}{h_1} \left( \frac{u_2 - u_1}{h_1} \right) - \lambda \rho_1 u_1 = -\frac{1}{h_1}
\]

and

\[
u_{k+1} = 0.
\]

Note that the first boundary condition (2.6) is consistent with the differential equation since it is the same as creating a dummy node \(u_0\), allowing \(j = 1\) in (2.5) and setting

\[-\hat{\gamma}_0 \frac{u_1 - u_0}{h_0} = 1.\]

Discrete analogues of \(\gamma\) and \(\rho\) equal \(\hat{\gamma}_j = \rho_j = i\omega\) in accordance with the definition (2.2) of their continuous counterparts on \([0, +\infty)\).

As follows from [11, 18], \(u_1 = -\frac{\hat{h}_1}{\gamma_1}\), where the so-called discrete impedance function \(f_k\) is the Stieltjes fraction or S-fraction

\[
f_k(\lambda) = \frac{1}{\hat{h}_1 \lambda + \frac{1}{\hat{h}_1} + \frac{1}{\hat{h}_2 \lambda + \cdots + \frac{1}{\hat{h}_k \lambda + \frac{1}{\hat{h}_k}}}}.
\]

It is easy to see that this S-fraction formulation is algebraically equivalent to the continued fraction formulation of [16] with a proper choice of the steps.

Overall, for the error functional we get

\[
\delta_k = \max_{\lambda \in [0,1]} s(\lambda) \left| \sqrt{\lambda} - \frac{1}{\gamma u_1} \right| = \max_{\lambda \in [0,1]} s(\lambda) \left| \sqrt{\lambda} - \frac{1}{f_k(\lambda)} \right|.
\]

We are now prepared to show that the proper choice of steps for the finite-difference discretization makes \(\delta_k\) attain the minimum value for all \(\delta^{\alpha,\beta}_k\).

3. Optimality. Let us return to the finite-difference PML for the general case given by (1.5) and show that the problem of its optimal finite-difference approximation can be reduced (possibly not uniquely) to the problem considered in the previous section, i.e.,

\[
\min_{\alpha,\beta} \delta^{\alpha,\beta}_k = \min_{\beta} \delta^0_k \equiv \min_{h,h} \delta_k.
\]

Fortunately, the latter is a well-studied problem of Chebyshev rational approximation.
In terms of the new variable $\bar{x}$ defined by (1.5) we obtain the same problem as (2.1) but with the following new coefficients:

$$\gamma = \frac{i\omega}{i\omega \alpha + \beta}, \quad \rho = -\omega^2 \left( \alpha + \frac{\beta}{i\omega} \right) \text{ if } \bar{x} > 0.$$

We use a finite-difference discretization similar to (2.5)–(2.7). Let us denote the discrete counterparts of $\alpha$ and $\beta$, respectively, by $\alpha_j, \hat{\alpha}_j$ and $\beta_j, \hat{\beta}_j$. Formally we assume that $\alpha_j, \beta_j$ and $\hat{\alpha}_j, \hat{\beta}_j$ reside at points $x_j$ and $\hat{x}_j$, respectively. Then the PML finite-difference solution satisfies (2.5)–(2.7) but with $\hat{\gamma}_j = \rho_j = 1$ and the new grid "steps"

$$a_j = h_j \left( \hat{\alpha}_j + \frac{\hat{\beta}_j}{i\omega} \right), \quad \hat{a}_j = -\omega^2 \hat{h}_j \left( \alpha_j + \frac{\beta_j}{i\omega} \right), \quad j = 1, \ldots, k. \tag{3.1}$$

The general finite-difference PML solution can be defined similarly to the one from the previous section if we use the new steps (3.1) in (2.5)–(2.7). We denote by $f_{\alpha,\beta}^k(\lambda, \omega)$ the discrete impedance function presented by formula (2.8) with the new steps (3.1). This function $f_{\alpha,\beta}^k$ is still a regular continued fraction but since its coefficients can be complex now, it is generally not an S-fraction anymore. We can equivalently rewrite the PML error (1.6) as

$$\delta_{\alpha,\beta}^k = \max_{\lambda \in [0,1], \omega \in [0,\omega_{\text{max}}]} s(\lambda) \left| \sqrt{\lambda} - \frac{1}{i\omega f_{\alpha,\beta}^k(\lambda, \omega)} \right|. \tag{3.2}$$

It is easy to check that when $\alpha = 0$ the dependence on $\omega$ of function $f_{\alpha,\beta}^0(\lambda, \omega)$ is such that $i\omega f_{\alpha,\beta}^0(\lambda, \omega) = f_{\beta}^0(\lambda)$. Obviously, for every set $\hat{\beta}_j, \beta_j$ we can find real steps $h_j, \hat{h}_j$, so that $f_{\beta}^0(\lambda) = f_{\beta}^k(\lambda)$, and hence the error functional (3.2) with $\alpha = 0$ coincides with (2.9), i.e.,

$$\delta_k = \delta_{0,\beta}^k.$$

The following proposition shows that the general complex transformations cannot produce better PML approximations than their purely imaginary counterparts.

**Proposition 3.1.** For any $\alpha$ and any $\beta \neq 0$ we have $\delta_{\alpha,\beta}^k \geq \delta_k$.

**Proof.** From the definitions it follows that

$$\left[ i\omega f_{\alpha,\beta}^k(\lambda, \omega) \right]_{\omega=0} = i\omega f_{\beta}^0(\lambda, \omega) = f_{\beta}^k(\lambda);$$

therefore,

$$\delta_{\alpha,\beta}^k = \max_{\lambda \in [0,1], \omega \in [0,\omega_{\text{max}}]} s(\lambda) \left| \sqrt{\lambda} - \frac{1}{i\omega f_{\alpha,\beta}^k(\lambda, \omega)} \right| \geq \max_{\lambda \in [0,1], \omega = 0} s(\lambda) \left| \sqrt{\lambda} - \frac{1}{i\omega f_{\beta}^0(\lambda, \omega)} \right| = \delta_{0,\beta}^k = \delta_k. \quad \square$$

The proposition allows us to reduce minimization of $\delta_{\alpha,\beta}^k$ to minimization of $\delta_k$, which is a well-studied problem of Chebyshev rational optimization. Using the fact...
that $\sqrt{\lambda}$ is hypernormal on the support $[a, b]$ of $s$ [23], we conclude that the existence and uniqueness theorems for the Chebyshev (optimal) rational approximation, which can be found in [2, Chapter II, Theorems 33 and 34], are applicable. Since the function $f_k$ is an S-fraction, it is a $[(k - 1)/k]$ real rational function (though the opposite statement is not true in general). However, we will be looking for the optimal approximation in the form of the $[(k - 1)/k]$ real rational function. Such an approximant is irreducible and has exactly $2k + 1$ alternating points (set of ordered noncoinciding points $\lambda_j$, where the weighted error is equal to $(-1)^j \delta_k$) on $[a, b]$. Hence, the $[(k - 1)/k]$ optimal rational approximant must have $2k$ interpolation points on $(a, b)$, so it must be a Markov–Stieltjes function [15]. But any Markov–Stieltjes function can be presented as an S-fraction [6, Theorem 5.1.2, Corollary 2], i.e., as (2.8) with all $h_j$ and $\hat{h}_j$ being positive. The above consideration can be summarized in the following proposition.

**Proposition 3.2.** Let the support of $s$ be a segment $[a, b] \subseteq [0, 1]$, and let $s$ be continuous and positive on $[a, b]$. Then there exists the unique minimum of $\delta_k$ with the optimal approximant being an S-fraction (2.8).

The optimal approximant can be uniquely characterized as the one that has exactly $2k + 1$ alternation points, all of which are located on $[a, b]$.

From the above proposition we obtain the following corollary.

**Corollary 3.3.** The choice of parameters $\alpha_j^{\text{opt}} = \hat{\alpha}_j^{\text{opt}} = 0$, $\beta_j^{\text{opt}} = h_j$, $\hat{\beta}_j^{\text{opt}} = \hat{h}_j$, which minimize $\delta_k$, defines a solution that minimizes the functional $\delta_{\alpha,\beta}^k$.

**Remark 3.1.** Generally, there is no uniqueness in optimal approximation by complex rational functions even in approximating real functions on real intervals [25, Chapter 5]: that is why we cannot say if our optimal solution is unique. Though, obviously, $\beta^{\text{opt}}$ is unique, so the remaining question is whether there exists an optimal pair $(\alpha^{\text{opt}}, \beta^{\text{opt}})$ with $\alpha^{\text{opt}} \neq 0$.

**Remark 3.2.** There is an important reason why we chose the $L^\infty$ norm in our “cost” functional, instead of the $L_2$ norm traditionally used in the literature on PML, e.g., in [10], in addition to the fact that the former provides more reliable bounds. Even real rational approximation problems in $L^p$, $1 \leq p < \infty$, may have more than one optimal solution [20, section 2.3], so using the $L^\infty$ norm at least provides us the uniqueness for $\beta^{\text{opt}}$.

**Remark 3.3.** An interesting question is if a high order or spectral Galerkin PML discretization can perform better than the optimal finite-difference scheme, and the answer is that it cannot. The reason is that a Galerkin (and, generally, any Galerkin–Petrov) process on any $k$-dimensional subspace also generates a $[(k - 1)/k]$ rational impedance [13], so it cannot do better than the optimal $[(k - 1)/k]$ approximation.

4. **Rational approximation and optimal grids.** Perhaps the simplest way to construct a rational approximation is to compute a Padé approximant satisfying the conditions

$$\frac{d^i}{d\lambda^i} \left[ f_k(\lambda) - \sqrt{\lambda} \right]_{\lambda=1} = 0, \quad i = 0, \ldots, 2k - 1,$$

in which case the PML becomes algebraically equivalent to the ABC by Engquist and Majda [14]. However, such an approximation is not efficient for small $\lambda$. Better results can be achieved using a more general continued fraction ABC based on multipoint Padé approximants [16], but the double root interpolation used there does not produce the alternation of the error and thus cannot arrive at the optimal approximation.

Real optimal rational approximations in rare cases can be obtained in a closed
form, and in most cases they are obtained numerically, but in all cases the algorithms are based on the alternation property which was stated in Proposition 3.2.

### 4.1. Zolotarjov’s approximation.

Consider first the case $\lambda_{\text{min}} > 0$. The interval $[\lambda_{\text{min}}, 1]$ can be linearly shifted onto $[1, 1/\lambda_{\text{min}}] = [1, 1/\kappa']$ with $\kappa' = \sqrt{\lambda_{\text{min}}}$. Let $\kappa = \sqrt{1 - \kappa^2}$. Zolotarjov found a $[(k - 1)/k]$ rational function $\tilde{r}$ such that

$$\left\| 1 - \sqrt{\lambda} \tilde{r}(\lambda) \right\|_{C[1, 1/\kappa']} = \inf \left\{ \left\| 1 - \sqrt{\lambda} r(\lambda) \right\|_{C[1, 1/\kappa']} \right\}.$$

When the error is small, this optimization problem is close to (1.6)–(1.7), in the sense that if the error of one of these approximation is equal to $\varepsilon$, then the error of the other is $\varepsilon + O(\varepsilon^2)$.

**Theorem 4.1** (see Zolotarjov, 1887 [20]). *The best approximant is given by*

$$\tilde{r}(\lambda) = D \prod_{l=1}^{k} \left( \frac{\lambda + c_2 l}{\prod_{l=1}^{k} (\lambda + c_2 l - 1)} \right),$$

where

$$c_l = \frac{\text{sn}^2 (lK/(2k); \kappa)}{\text{cn}^2 (lK/(2k); \kappa)} \quad l = 1, \ldots, 2k - 1,$$

$K = K(\kappa)$ is the complete elliptic integral, and the number $D$ is uniquely determined by the condition

$$\max_{C[1, 1/\kappa']} \left[ 1 - \sqrt{\lambda} \tilde{r}(\lambda) \right] = -\min_{C[1, 1/\kappa']} \left[ 1 - \sqrt{\lambda} \tilde{r}(\lambda) \right].$$

*Remark 4.1.* If $\lambda_{\text{min}}$ is small, then $\kappa$ is close to 1. Standard subroutines for computing elliptic functions fail at the very beginning, because they accept $\kappa$ and compute $\kappa' = \sqrt{1 - \kappa^2}$, losing significant digits. We recommend computing the elliptic functions by the arithmetic-geometric mean method [1, Chapters 16 and 17] in terms of $\kappa' = \sqrt{\lambda_{\text{min}}}$. Recall that the asymptotic convergence factor is given by formula (1.9).

In Figure 1 one can see the Chebyshev alternation of the error via $\lambda$ on the optimal interval for a Zolotarjov’s grid. The $L^\infty$ norm of the error for $\lambda_{\text{min}} = 10^{-3}$ is given in Figure 2; it is in good agreement with the estimate (1.9).

### 4.2. Approximation on $[0, 1]$.

In the case $\lambda_{\text{min}} = 0$ we numerically solved the optimization problem (1.6), (1.8). Here we followed the solution of this problem given in [24] with some minor modifications.

For small values of $k$ we used the unconditionally converging differential correction method [20, section 2.5], the convergence of which, however, deteriorates rapidly with the increase of $k$. To solve the associated convex nonsmooth optimization problem, we exploited the Fortran 90 package SOLVOPT by A. Kuntsevich and F. Kappel (see http://bedvgm.kfunigraz.ac.at:8001/alex/solvopt).

For larger $k$ the Remez method [20, section 2.5] was exploited. This algorithm converges quadratically, provided a good initial guess is given, but otherwise it may diverge. Therefore, we implemented an extrapolation procedure in the spirit of [24] to obtain good initial iterants. The Fortran 90 multiple precision package [5] was incorporated into our Fortran program realizing the Remez method.
An error estimate is presented by formula (1.10); it follows from [21]. The actual distribution of the error for $k = 10$ is plotted in Figure 3; similarly to the Zolotarjov’s error it exhibits Chebyshev alternating properties according to Proposition 3.2. The error graph in Figure 4 is in very good agreement with the error estimate (1.10).

4.3. Computation of grid steps. The optimal rational approximations obtained above can be represented in terms of poles and residues as

$$f_k(\lambda) = \sum_{i=1}^{k} \frac{y_i}{\lambda - \theta_i},$$

but to construct the finite-difference scheme we need them in the form of an S-fraction (2.8). A recursive algorithm of computing $h_i$ and $\hat{h}_i$ from $y_i$ and $\theta_i$ based on the Lanczos method is given in [11, subsection 3.1]. In Figure 5 we show grids for Zolotarjov approximation on $[0.001, 1]$- and $[0, 1]$-optimal approximations. Qualitatively, both the grids behave similarly to other optimal or Gaussian grids described in [11, 12, 3, 18, 13]; i.e., they exhibit gradual refinement towards the origin and alternation of primary and dual nodes. Specifically, both grids are close to geometric progressions, which corresponds well to the asymptotic property of optimal grids on spectral intervals with large condition numbers, discovered in [18]. Since the $[0, 1]$-
The computed impedance error of Zolotarjov’s grids is designed to absorb at all incidence angles, it requires thicker PML than the Zolotarjov grid.

4.4. Comparison with standard PML. The first advantage of the new method of implementation of PML over the standard implementation is that the errors of the newly computed meshes depend only on the incidence angle $\theta = \arccos(\sqrt{\lambda})$ and not on the frequency. The practical importance of this fact is that the user can fix the mesh only once, depending on the requirements of the accuracy and the spectral range in question. Of course, the reflection coefficient of such a mesh will also depend on the discretization of the interior domain, and thus the total error of the ABC will not be better than the error of the interior discretization.

The second important property is that these meshes give an optimal or near-optimal error for a chosen spectral interval, and no ad hoc or numerical optimization of parameters is necessary after the mesh is defined.

Previously, in the literature devoted to PML absorbing conditions, near-optimal solutions for the parameters were obtained by various numerical optimization schemes [10, 8]. Comparing the magnitudes of these errors to the errors of the new optimal PML, we conclude that the new scheme produces a much smaller error. From Figure 2 we see that a 5 point Zolotarjov’s mesh gives an error of approximately $10^{-4}$ for all $\lambda \in [10^{-3}, 1]$, which corresponds to the incident angles $\theta \in [0, 88^\circ]$, and a 10 point grid...
mesh gives an error of $10^{-8}$ for the same angles, all independent of the frequency. For comparison, the PML discretizations in [10] for approximately the same range of the angles give an error that varies (depending on the frequency) between $10^{-2}$ and $10^{-1}$ (5 point mesh), and between $10^{-3}$ and $10^{-2}$ (10 point mesh).

Among the disadvantages of the proposed method we see the nontrivial discretization of the time-domain problem.

5. Time-domain realization. The S-fraction representation of the impedance function $f_k$ provides stability of the time-domain solution. The poles of S-fractions are real negative and their residues are real positive, which is sufficient for the absence of exponentially growing modes [22].

Unfortunately, our modification of the equation in the absorbing region makes impossible the simple variable-splitting time-domain realization that is commonly used in the PML methods. Moreover, the equation in the absorbing region becomes noncausal, and hence the discretized system has to be implicit.

We will first consider the discretization of our two-dimensional wave equation with the absorbing region in the half-space $x > 0$; the case with absorbing layers on the four edges of a square differs only by the treatment of the corners; i.e., the subdomains in which the imaginary stretching is applied to both coordinates. The treatment of the two-dimensional domains will be briefly discussed in the end of the

Fig. 3. The impedance error $\sqrt{\lambda} - \tilde{r}(\lambda)$ of optimal grid for the segment $[0, 1]$ and $k = 10$ as a function of $\lambda$. 
Let us first Fourier transform (2.1)–(2.2) back to the two-dimensional wave problem while again keeping the same name for the function:

\[ u_{xx} + u_{yy} - u_{tt} = 0 \text{ if } x < 0, \quad u_{xxtt} + u_{yy} - u_{tt} = 0 \text{ otherwise,} \]

with the interface conjugation conditions

\[ u(0-, y, t) = u(0+, y, t), \quad u(0-, y, t)_x = u(0+, y, t)_{xt} \]

and the boundary condition \( u(+\infty, y, t) = 0 \). We will demonstrate the discretization in time and in the \( x \)-direction, while leaving the problem continuous in the \( y \)-direction.

For \( x < 0 \) we write the discretization of the equations in the standard way:

\[
\begin{align*}
\frac{d}{dt}u_j^- &= \frac{V_j^- - V_{j-1}^-}{h_j^-} + W_j^- y, \\
\frac{d}{dt}V_j^- &= \frac{u_{j+1}^- - u_j^-}{h_j^-}, \quad u_{k+1}^- = 0, \\
\frac{d}{dt}W_j^- &= u_{j,y}^-.
\end{align*}
\]

\( j = 1, \ldots, k, \)
Here the unknowns are numbered from right to left with $u_{-1}^-$ and $V_0^-$ being placed at the boundary of the PML region. We note that the right-to-left numbering of the unknowns implies that the steps $h_j^-$, $\hat{h}_j^-$ are negative.

The only change in the PML region, i.e., for $x > 0$, is that the steps $h_j^+$ and $\hat{h}_j^+$ get divided by $i\omega$, which leads to the following system in time domain:

$$\frac{d}{dt}u_j^+ = \frac{d}{dt}V_j^+ = \frac{d}{dt}W_j^+, \quad j = 1, \ldots, k,$$

$$\frac{d}{dt}V_j^+ = \frac{d}{dt}u_{j+1}^+ - \frac{d}{dt}u_j^+ + \frac{u_{j+1}^+ - u_j^+}{h_j^+}, \quad u_{k+1}^+ = 0, \quad j = 1, \ldots, k,$$

$$\frac{d}{dt}W_j^+ = u_{j,y+1}^+ - u_{j,y}^+, \quad j = 1, \ldots, k.$$

Here the numbering of the unknowns is left to right; hence the steps $h_j^+$, $\hat{h}_j^+$ are positive.

It is clear that to get the total number of equations equal to the total number of unknowns we need two extra equations; these are

$$u_1^- = u_1^+, \quad V_0^- = V_0^+.$$

The time discretization follows naturally: we place variables $u_j^-$ at time levels $n\Delta t$ and variables $V_j^-$ and $W_j^-$ at time levels $(n + 1/2)\Delta t$, where $\Delta t > 0$ is a time step. The system (5.1)–(5.3) is thus solved in the standard explicit “leap-frog” fashion.
For $x > 0$, however, we have to place variables $u_j^+$ and $V_j^+$ at the same time levels $n\Delta t$, and only the variables $W_j^+$ at time levels $(n+1/2)\Delta t$; thus (5.4), (5.5), and the second equation in (5.7) will compose the system of equations that needs to be solved at every time step. With the use of the notation

$$[D_t f]^{m+1/2} = \frac{f^{m+1} - f^m}{\Delta t}, \quad [A_t f]^{m+1/2} = \frac{f^{m+1} + f^m}{2},$$

where one can put $m = n$ or $m = n + 1/2$, we rewrite the scheme (5.1)–(5.7) as

$$[D_t u_1]^n + 1/2 = \frac{V_1^n - u_1^n}{h_1} + W_1^n + 1/2,$$

$$[D_t u_j]^n + 1/2 = \frac{V_j^n - u_j^n}{h_j} + W_j^n + 1/2, \quad j = 2, \ldots, k,$$

$$[D_t V_j]^n = \frac{u_{j+1}^n - u_j^n}{h_j}, \quad j = 1, \ldots, k,$$

$$[D_t W_j]^n = u_{j,y}^n, \quad j = 1, \ldots, k,$$

for $x \leq 0$, and

$$[D_t u_j]^n + 1/2 = \frac{[D_t V_j]^n + 1/2 - [D_t V_{j-1}]^n + 1/2}{h_j} + W_{j,y}^n + 1/2, \quad j = 1, \ldots, k,$$

$$[D_t V_j]^n + 1/2 = \frac{[D_t u_{j+1}]^n + 1/2 - [D_t u_j]^n + 1/2}{h_j}, \quad j = 1, \ldots, k,$$

$$[D_t W_j]^n = u_{j,y}^n, \quad j = 1, \ldots, k,$$

for $x > 0$. We note that variables $V_j^+$ with $j > 0$ can be excluded from the scheme analytically by substituting the second of the three equations above into the first. The time marching follows in an obvious fashion.

The discretization of the two-dimensional model is performed as a tensor product of the two one-dimensional discretizations. It can be seen that in the corners of a rectangular domain in this case the equation is independent of time. In fact, it turns out that the Helmholtz equation $u - u_{xx} - u_{yy} = 0$ is solved there, with the time dependent boundary conditions. For such an equation on a given mesh one can compute the Neumann-to-Dirichlet map on the boundary and include only the variables defined on the boundary in the total scheme.

It becomes clear that the computational cost of the new boundary condition amounts to solving a linear system with a tridiagonal matrix of dimension $k$ for each point on the interface at every time step. This cost will be linear in $k$ with the constant close to the one for the implementation of standard Berenger’s PML with variable splitting. The computation of the corner regions for a two-dimensional problem requires computing at each time step a partial Neumann-to-Dirichlet map on a $k \times k$ square, which again is an operation proportional to the total number of nodes in the absorbing region.

### 6. Numerical experiments.

We performed a series of experiments for the following elongated model: the source and the receiver, positioned on the same vertical
Fig. 6. The relative error for meshes A and B of Experiment 1.

The relative errors (i.e., with $s(\lambda)$ defined as in (1.7)—notice that mesh B was designed to provide minimum error in a different norm!) for these meshes are presented in Figure 6, and we see that these two meshes have approximately the same order of
approximation on the spectral interval of interest. It is thus not surprising that the $L^\infty(t)$ errors that we obtain are similar too, as seen in the line $F_{\text{max}} = 1$ of Table 1.

It is clear that the error is partly due to the reflection of the evanescent modes from the boundary between the interior propagative part of the domain and the absorbing part. To consider the importance of evanescent reflections, we increase the frequency of the source, thus keeping the incidence angles of the experiment unchanged, while effectively increasing the distance from the absorbing boundary to the receiver in terms of the wavelengths, and decreasing the effect of the evanescent reflections. The errors resulting from this change (with frequencies $F_{\text{max}} = 2$ and $F_{\text{max}} = 4$) are shown on the corresponding lines of Table 1. We note that with this change the boundary conditions become more effective for all values of the discretization of the interior for both meshes, but the effect of the decrease of the evanescent modes is especially notable for fine discretization of 32 points per wavelength, at which the boundary conditions approach the saturation error.

For the ABCs to be most effective, they need to be able to absorb not only the propagative modes of the solution but also the evanescent modes. Our proposed PML ABCs are optimal in terms of the absorption of the propagative part; however, the evanescent modes were left untreated. The obvious solution to this problem is to place this boundary condition at a distance to the closest receiver; this will ensure that the evanescent modes of the solution get absorbed. However, in the problems in which the receiver stays closer to the boundary conditions than any scatterer, the necessity of placing the boundary conditions away from the receiver clearly decreases the effectiveness of such conditions. This problem can be cured by using the regular nonabsorbing optimal meshes [3, 11, 12, etc.] in the layers between the receiver and the boundary conditions. This approach allows one to solve the problem of absorbing both the evanescent and propagative modes in an optimal way.

It can be seen from the results of these experiments that in the cases in which the spectral interval is known a priori (as above) it is reasonable to use Zolotarjov’s meshes, specifically designed for such an interval. However, when such a mesh is not available, or in the case where the spectral interval is not known in advance (such as in models with multiple reflections), we expect that it will be advantageous to use the universal optimal mesh of the same type as “B” above, which satisfies the accuracy requirements.

7. Conclusions. We have shown that the optimal choice of the attenuation parameters for Berenger PML ABCs can be achieved in the limiting case of the purely imaginary coordinate stretching approach. Indeed, it was questioned by one of the reviewers if the proposed approach can and should be included in the class of PML boundary conditions, because, unlike those conditions, it leads to a diffusion-type equation in the absorbing region. However, the situation when a minimum is attained on the boundary of the closure of a set is often encountered in optimization theory, and thus our new ABC is a part of the closure of the classical PML set.
For this limiting case, the problem becomes frequency independent and can be viewed as a classical Chebyshev rational approximation of the square root function, and the parameters that need to be chosen are the steps of the grid of the resulting finite-difference scheme. We consider two different error functions and show how to obtain the optimal grid steps for them. The resulting scheme exhibits attenuative properties superior to those of the classical PML known to the authors; the goal attenuation of 1% can be achieved using as few as five points in the absorbing region. The drawback of the resulting scheme is that, unlike the classic PML, it needs to be implemented in a fashion that is different in the absorbing and the propagating regions. The details of the discretization of a two-dimensional scalar wave problem are discussed and the results of numerical experiments are presented.

Acknowledgments. The authors are grateful to R. Varga for useful discussions and to D. Bailey (NASA Ames Research Center) for the publication of package [5] on the Internet.

REFERENCES

[19] P. G. Petropoulos, Reflectionless sponge layers as absorbing boundary conditions for the


