



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Journal of Computational Physics 210 (2005) 183–224

JOURNAL OF
COMPUTATIONAL
PHYSICS

www.elsevier.com/locate/jcp

Numerical solution of the nonlinear Helmholtz equation using nonorthogonal expansions

G. Fibich^{a,1}, S. Tsynkov^{b,*}

^a *Department of Applied Mathematics, School of Mathematical Sciences, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel*

^b *Department of Mathematics and Center for Research in Scientific Computation (CRSC), North Carolina State University, Box 8205, Raleigh, NC 27695, USA*

Received 18 November 2004; received in revised form 7 April 2005; accepted 7 April 2005
Available online 17 June 2005

Abstract

In [J. Comput. Phys. 171 (2001) 632–677] we developed a fourth-order numerical method for solving the nonlinear Helmholtz equation which governs the propagation of time-harmonic laser beams in media with a Kerr-type nonlinearity. A key element of the algorithm was a new nonlocal two-way artificial boundary condition (ABC), set in the direction of beam propagation. This two-way ABC provided for reflectionless propagation of the outgoing waves while also fully transmitting the given incoming beam at the boundaries of the computational domain. Altogether, the algorithm of [J. Comput. Phys. 171 (2001) 632–677] has allowed for a direct simulation of nonlinear self-focusing without neglecting nonparaxial effects and backscattering. To the best of our knowledge, this capacity has never been achieved previously in nonlinear optics.

In the current paper, we propose an improved version of the algorithm. The principal innovation is that instead of using the Dirichlet boundary conditions in the direction orthogonal to that of the laser beam propagation, we now introduce Sommerfeld-type local radiation boundary conditions, which are constructed directly in the discrete framework. Numerically, implementation of the Sommerfeld conditions requires evaluation of eigenvalues and eigenvectors for a non-Hermitian matrix. Subsequently, the separation of variables, which is a key building block of the aforementioned nonlocal ABC, is implemented through an expansion with respect to the nonorthogonal basis of the eigenvectors. Numerical simulations show that the new algorithm offers a considerable improvement in its numerical performance, as well as in the range of physical phenomena that it is capable of simulating.

© 2005 Elsevier Inc. All rights reserved.

* Corresponding author. Tel.: +1 919 515 1877; fax: +1 919 515 3798.

E-mail addresses: fibich@math.tau.ac.il (G. Fibich), tsynkov@math.ncsu.edu (S. Tsynkov).

URLs: <http://www.math.tau.ac.il/~fibich> (G. Fibich), <http://www.math.ncsu.edu/~stsynkov> (S. Tsynkov).

¹ The work of the first author was supported by Grant No. 2000311 from the United States – Israel Binational Science Foundation (BSF), Jerusalem, Israel.

Keywords: Kerr media; Intense laser light; Diffraction; Nonparaxiality; Nonlinear self-focusing; Backscattering; Critical and subcritical nonlinearity; Solitary waves; Counter-propagation; Fourth-order approximation; Iterative solution; Separation of variables; Nonlocal artificial boundary conditions (ABCs); Sommerfeld radiation boundary conditions

1. Introduction

1.1. Background

The objective of this paper is to introduce, implement, and test a new and improved numerical algorithm for studying nonlinear self-focusing of time-harmonic electromagnetic waves. The phenomenon of self-focusing takes place in optical propagation through many “conventional” media, such as water, silica, or air, provided that the intensity of the impinging electromagnetic wave is sufficiently high. The physical mechanism that leads to self-focusing is known as the Kerr effect. At the microscopic level, the Kerr effect may originate from electrostriction, nonresonant electrons, or from molecular orientation. At the macroscopic level the Kerr effect is manifested through an increase in the index of refraction, which is proportional to the intensity of the electric field $|E|^2$. Since light rays bend toward regions with higher index of refraction, an impinging laser beam would become narrower as it propagates, a phenomenon known as self-focusing. For more information on self-focusing, see, e.g., [1–3].

Since nonlinear self focusing leads to nonuniformities in the refraction index, a part of the incoming beam gets reflected backwards, a phenomena known as nonlinear backscattering. At present, very little is known about it, except for the general belief that it is “small.” Since, however, small-magnitude mechanisms can have a large effect in nonlinear self-focusing in bulk medium [3], one needs to be able to accurately quantify the magnitude of backscattering and study how it may affect the beam propagation. From the standpoint of applications in modern science and engineering, the capability to quantitatively analyze and predict the phenomena of nonlinear self-focusing and backscattering is extremely important for a large number of problems. Those range from remote atmosphere sensing (when an earth-based powerful laser sends pulses to the sky [4], and the backscattered radiation accounts for the detected signal), to laser surgery (propagation of laser beams in tissues), to transmitting information along optical fibers. There are other possible applications, e.g., all-optical switching in electrical/electronic circuits, that involve interactions (collisions) between co-propagating or counter-propagating beams. It is well known that within the framework of a simpler model based on the nonlinear Schrödinger equation, these beams have the form of spatial solitary waves (solitons), and that collisions between solitons are elastic (i.e., involve no power losses). However, whether such collisions remain elastic in more comprehensive models, such as the nonlinear Helmholtz question, is currently an open problem.

1.2. The nonlinear Helmholtz equation

The simplest theoretical (and experimental) setup for the propagation of electromagnetic waves in Kerr media is shown in Fig. 1. An incoming laser beam with known characteristics impinges normally on the planar interface $z = 0$ between the linear and nonlinear medium. The time-harmonic electric field $E = E(\mathbf{x}, z)$ is assumed to be linearly polarized and is governed by the scalar nonlinear Helmholtz equation (NLH):

$$\begin{aligned} (\partial_{zz} + \Delta_{\perp})E + k^2 E &= 0, \\ k^2 &= k_0^2(1 + \epsilon|E|^{2\sigma}), \quad \sigma > 0, \quad \mathbf{x} \in \mathbb{R}^{D-1}, \quad z \geq 0, \end{aligned} \tag{1}$$

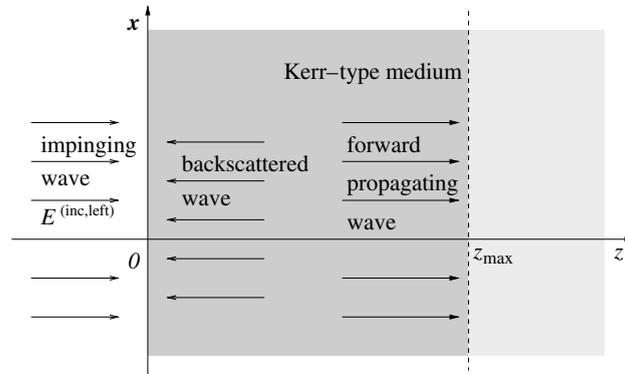


Fig. 1. Schematic for the propagation of waves in Kerr media.

where k_0 is the linear wavenumber, $\epsilon = 2n_2/n_0$, n_0 is the linear index of refraction, n_2 is the Kerr coefficient, and $\Delta_{\perp} = \partial_{x_1 x_1} + \dots + \partial_{x_{D-1} x_{D-1}}$ is the $D - 1$ dimensional transverse Laplacian (the diffraction term). The value of σ is equal to one for the physical Kerr effect. However, for reasons that will become apparent in Section 1.5, other values of σ are also considered in this study. Note also that in all the physical problems mentioned earlier, the nonlinearity in (1) is typically weak, i.e., $\epsilon |E^{(inc,left)}(\mathbf{x})|^{2\sigma} \ll 1$, where $E^{(inc,left)}$ is the impinging wave, see Fig. 1. The reason for this is that the physical value of the Kerr index n_2 is so small, that even for high-power lasers the nonlinear change in the index of refraction is usually small compared with the linear index of refraction, i.e., $n_2 |E|^2 \ll n_0$.

The nonlinear medium occupies the half-space $z \geq 0$. Consequently, the NLH (1) should be supplemented with boundary conditions at $z = 0$ and at $z = +\infty$. Although global existence of its solutions is currently an open problem, there are various indications that, at least in some cases, the NLH is solvable (see Section 1.5). In fact, one of our primary long term goals in this study is to address the solvability issue with the help of our numerical methodology. Assuming that the solution does exist globally, for large propagation distances (i.e., as $z \rightarrow +\infty$) it can either diffract, in which case the propagation becomes linear, or it can maintain its shape through a balance of the focusing nonlinearity and diffraction (i.e., converge to a soliton). In either case, as $z \rightarrow +\infty$, E will have no left-propagating components, i.e., E will only be composed of the right-traveling waves. Since the actual numerical simulation is carried out on a truncated domain $0 \leq z \leq z_{max}$, the desired behavior of the solution as $z \rightarrow +\infty$ has to be captured by a far-field artificial boundary condition (ABC) at the artificial boundary $z = z_{max}$. This boundary condition should guarantee the reflectionless propagation of all the waves traveling toward $z = +\infty$ through the interface $z = z_{max}$. Often, boundary conditions designed to ensure the transparency of the outer boundary to the outgoing waves are referred to as radiation boundary conditions [5].

The situation at the interface $z = 0$ is more complicated, as the total field there, $E(\mathbf{x}, 0)$, is composed of a given incoming component $E^{(inc,left)}(\mathbf{x})$ and the unknown backscattered (i.e., outgoing) component $E^{(scat)}(\mathbf{x}, 0)$, i.e.,

$$E(\mathbf{x}, 0) = E^{(inc,left)}(\mathbf{x}) + E^{(scat)}(\mathbf{x}, 0). \tag{2}$$

As such, the boundary condition at $z = 0$ has to provide for the reflectionless propagation of any outgoing (left-traveling) wave through this interface, and at the same time has to be able to correctly prescribe the incoming (right-traveling) field. A two-way ABC that possesses the required capabilities has been first implemented in [6].

In addition to the foregoing simplest setup, we can also use our algorithm to solve more elaborate problems. For example, instead of assuming that the nonlinear Kerr medium occupies the entire

semi-space $z \geq 0$ and as such, that it can only be excited by waves coming from the left, we can consider the Kerr medium to be confined to the “slab” $0 \leq z \leq z_{\max}$, as is frequently the case in experiments. Then, the excitation can be provided by waves impinging on both interfaces $z = 0$ and $z = z_{\max}$. As such, from the standpoint of setting the two-way ABCs these interfaces become equal – either has to be transparent for its corresponding outgoing waves while simultaneously being able to fully transmit the corresponding incoming field. The latter setup will be instrumental for simulations of coherent counter-propagating beams.

1.3. Transverse boundary conditions

Any of the foregoing formulations requires setting boundary conditions in the transverse direction(s) $x_j, j = 1, \dots, D - 1$. Although in some cases the domain filled by the Kerr medium extends all the way to infinity in the transverse direction(s), for the purpose of practical computing, we truncate this domain at an external artificial boundary $\partial\Omega \times \mathbb{R}^+$ and set the boundary conditions for all (\mathbf{x}, z) such that $\mathbf{x} \in \partial\Omega \subset \mathbb{R}^{D-1}$ and $z \geq 0$. In our previous work [6–8], we assumed that the electric field vanished at this lateral artificial boundary: $E(\mathbf{x}, z)|_{\mathbf{x} \in \partial\Omega, z \geq 0} = 0$. While the approach in [6–8] enabled us, apparently for the first time ever, to solve the NLH as a genuine boundary value problem, there have still been indications that it could be improved. Indeed, the Dirichlet boundary conditions, while being easy to implement, possess the non-physical property of reflecting the waves back from the lateral boundaries into the computational domain. Even though the solution with the Dirichlet transverse boundary conditions can be expected to approximate well the original infinite-domain solution in the domain of physical interest, namely in the central region of the computational domain,² we still had to position the lateral boundaries sufficiently far away from the center so that to alleviate the undesirable reflections, see [6–8]. In addition, the Dirichlet boundary conditions have been found to sometimes produce spurious numerical resonances that prevented the algorithm of [6] from converging. Therefore, in the current paper we replace the Dirichlet boundary condition with a local Sommerfeld-type radiation boundary condition. Implementation of these Sommerfeld radiation boundary conditions at the transverse boundaries of the computational domain is the key algorithm improvement that we present in the current study, compared with [6]. As we shall see, the present methodology facilitates major gains in performance, allowing us to reduce the transverse width of the computational domain considerably. Moreover, it completely removes the aforementioned resonance problem.

1.4. Paraxial approximation and the nonlinear Schrödinger equation

Let r_0 be the initial radius of the impinging laser beam. We first introduce the dimensionless quantities $\tilde{\mathbf{x}}$, \tilde{z} , and A :

$$\tilde{\mathbf{x}} = \frac{\mathbf{x}}{r_0}, \quad \tilde{z} = \frac{z}{2L_{\text{DF}}}, \quad E = e^{ik_0z} (\epsilon r_0^2 k_0^2)^{-1/2\sigma} A(\mathbf{x}, z), \quad (3)$$

where $L_{\text{DF}} = k_0 r_0^2$ is the diffraction length. Then, by dropping the tildes we obtain from (1):

$$iA_z + \Delta_{\perp} A + |A|^{2\sigma} A = -4f^2 A_{zz}, \quad (4)$$

where $f = 1/r_0 k_0 \ll 1$ is the small nonparaxiality parameter.

² More precisely, as the location of the lateral boundaries approaches infinity, the corresponding solutions converge to the infinite-domain solution on any region compactly supported in the transverse direction, see [9] for more detail on the concept of convergence on bounded sub-domains.

The standard derivation of the nonlinear Schrödinger equation (NLS) is based on the assumption that the envelope A varies slowly. Then, one can neglect the right-hand side of (4) (i.e., apply the paraxial approximation) and obtain the NLS:

$$iA_z + \Delta_{\perp} A + |A|^{2\sigma} A = 0. \quad (5)$$

The NLS (5) is an evolution equation in which the variable z plays the role of “time.” Hence, it only needs to be supplemented by the initial condition at $z = 0$:

$$A(\mathbf{x}, 0) = (\epsilon r_0^2 k_0^2)^{1/2\sigma} E^{(\text{inc.}, \text{left})}(\mathbf{x}).$$

Subsequently, the NLS (5) is to be integrated by a “time”-marching algorithm for $z > 0$. We note that almost all self-focusing studies available in the Literature have used NLS-based models, rather than the NLH, the main reason being that the NLH is much a “tougher” object for analysis and for simulations. However, the NLH provides a more comprehensive physical model for self-focusing. In particular, it takes into account the effect of nonlinear backscattering, which the NLS model disregards. Indeed, once (5) has been integrated, the overall solution, according to (3), is the slowly varying amplitude A times the forward propagating oscillatory component $e^{ik_0 z}$.

The capability to account for the effects of nonparaxiality and backscattering makes the NLH model (1) more in-depth compared to the simpler NLS model (5). It is precisely the role of these two important effects in nonlinear wave propagation that the numerical methodology developed in this paper would enable us to focus on. In the overall perspective, however, the scalar NLH (1) itself does not represent the most general setup either. A yet more comprehensive model would be the vector NLH that also accounts for the vectorial nature of the electric field while still taking care of the phenomena of nonparaxiality and backscattering. Note that the vectorial effects, nonparaxiality, and backscattering are all of the same order of magnitude; see [10] for more detail.

1.5. Solitons and collapse

It is well known that solutions of the NLS (5) exist globally when $\sigma(D - 1) < 2$, the subcritical NLS, but can become singular (collapse) at finite propagation distances, when either $\sigma(D - 1) > 2$, the supercritical NLS, or when $\sigma(D - 1) = 2$, the critical NLS [2]. As shown by Weinstein [11], a necessary condition for singularity formation in the critical NLS is that the input beam power (i.e., $\|A(\cdot, 0)\|_2^2$) exceeds the critical power N_c . The value of N_c is equal to the power of the ground-state solitary wave solution of the NLS; this value can be calculated explicitly for $D = 2$ and numerically for $D > 2$.

In our subsequent numerical simulations in Section 4 we will focus on two cases. The first one is $D = 2$ and $\sigma = 2$ (quintic nonlinearity), for which the NLS (5) is critical. Although the actual physical setting for the critical NLS is $D = 3$ and $\sigma = 1$, which corresponds to propagation of waves in a bulk Kerr medium, one can expect the role of nonparaxiality and backscattering to be quite similar for both cases. This expectation should, in particular, pertain to the question, which has been open for many years, of whether the more comprehensive NLH model for nonlinear self-focusing eliminates the singular behavior that characterizes collapsing solutions of the critical NLS. Unfortunately, the fundamental issue of solvability of the NLH and regularity of its solutions still remains largely unaddressed. There have been indications that solutions to the NLH may exist even when the corresponding NLS solutions become singular, based on both numerical study of “modified” NLS equations [12–14], and on asymptotic analysis [15], but these studies did not take the backscattering effects into account. An additional argument in this direction is provided in our recent work [8], where we have used the numerical algorithm of [6], that does account for backscattering, to show the existence of initial conditions for which the solutions of the linearly damped NLS become singular, whereas the corresponding solutions of the linearly-damped NLH exist globally. In a very recent work

[16], Sever employed a Palais-Smale type argument and has shown that the NLH is solvable in the sense of H^1 and that the solution is not unique. His argument, however, only applies to self-adjoint operators, which, as we will see, is not the case in many physics-motivated situations. We therefore expect that the original numerical methodology of [6], its extension proposed in the current paper, as well as possible subsequent improvements, may eventually help us answer the outstanding question of global existence in the critical NLH.

The second case that we will consider is $D = 2$ and $\sigma = 1$ (cubic nonlinearity), which physically corresponds to nonlinear propagation in a planar waveguide (i.e., there is only one transverse dimension). In this subcritical case solutions to the NLS do not collapse. Instead, the laser beam can propagate in the Kerr medium over very long distances without changing its profile – the type of behavior often referred to as spatial soliton. In the past, solitons have primarily been studied as solutions to the NLS. Although it is generally expected that in the subcritical case the NLH will have similarly looking solutions, until now it was not actually possible to study the effect of nonparaxiality and backscattering on solitons. Even more so, the NLH appears particularly well suited for simulating interactions between counter-propagating solitons. Indeed, in the NLH framework the counter-propagating case can be naturally formulated as a boundary-value problem on the slab $0 \leq z \leq z_{\max}$, see Fig. 1, whereas in the NLS framework the two counter-propagating solitons will imply two opposite directions of marching.³

1.6. Brief contents of the paper

In Section 2, we use a continuous formulation to describe the algorithm for solving the NLH (1). The nonlinearity in Eq. (1) is addressed iteratively, so that only the standard constant-coefficient Helmholtz operator needs to be repeatedly inverted in the course of the iterations. This inversion is done using the separation of variables; the latter is rendered by expansion with respect to the eigenfunctions of the transverse Laplacian subject to the Sommerfeld boundary conditions at the lateral artificial boundaries. Besides its simplicity, the key reason for solving the linear constant coefficient Helmholtz equation via the separation of variables is that this approach facilitates setting the highly accurate nonlocal two-way ABCs at the boundaries $z = 0$ and $z = z_{\max}$. The key distinction between the method proposed hereafter and the previously developed technique of [6], is that in [6] the boundary condition at the lateral artificial boundary was zero Dirichlet and we could therefore use the conventional trigonometric Fourier expansion in the transverse direction. In contradistinction to that, hereafter we employ the Sommerfeld radiation boundary conditions in the transverse direction that lead to a non-Hermitian operator with nonorthogonal set of eigenfunctions. Some analytic aspects of this unconventional approach are derived. In particular, it is shown that the introduction of transverse radiation results in linear “damping” in the direction of propagation.

In Section 3, we describe the fourth-order discrete formulation of the algorithm: The finite-difference scheme that we use, the discrete local Sommerfeld boundary conditions at the lateral artificial boundaries, and the discrete nonlocal ABCs at $z = 0$ and $z = z_{\max}$. Also in this section, we discuss the expansion with respect to the nonorthogonal discrete eigenfunctions in the transverse direction.

In Section 4, we present computational results obtained with the new NLH solver for the critical NLH (collapsing solutions) and for the subcritical NLH (soliton propagation and counter-propagating solitons). The foremost observation to be made is that the introduction of the transverse Sommerfeld radiation boundary conditions offers major gains in numerical performance compared with the original methodology of [6]; these gains stem primarily from the possibility to substantially reduce the transverse size of the

³ Counter-propagating beams have been simulated using two coupled NLS equations [17], but this approach involves some approximations which are not needed within the NLH framework.

computational domain, see Section 4.1. In addition, the new approach allows us to improve the overall stability and robustness of the numerical algorithm.

Section 5 concludes the presentation with a discussion.

2. Algorithm – continuous formulation

For clarity of presentation, we first outline the algorithm using a continuous formulation. We consider from now on the case $D = 2$ that corresponds to propagation in a planar waveguide. Therefore, Eq. (1) becomes

$$\begin{aligned} (\partial_{zz} + \partial_{xx})E(x, z) + k^2 E &= 0, \\ k^2 &= k_0^2(1 + \epsilon|E|^{2\sigma}), \quad \sigma > 0, \quad x \in \mathbb{R}, \quad z \geq 0. \end{aligned} \quad (6)$$

Compared with the case $D = 3$, the two-dimensional setup offers a considerable reduction of the computational effort while preserving all the essential physical and numerical effects. For simplicity and with no substantial loss of generality, for all our numerical experiments we will choose the rectangular computational domain $\{-x_{\max} \leq x \leq x_{\max}, 0 \leq z \leq z_{\max}\}$. At the “upstream” and “downstream” boundaries $z = 0$ and $z = z_{\max}$ we will set the nonlocal two-way ABCs. At the transverse boundaries $x = \pm x_{\max}$ we will set local radiation boundary conditions of the Sommerfeld type.

2.1. Iteration scheme

We solve the NLH (6) iteratively by freezing the nonlinearity and reducing the NLH to a linear variable-coefficient equation on every iteration

$$\Delta E^{(j)} + k_0^2(1 + \epsilon|E^{(j-1)}|^{2\sigma})E^{(j)} = 0, \quad j = 1, 2, \dots, \quad (7)$$

where $\Delta = \partial_{zz} + \partial_{xx}$. The initial guess is typically chosen as $E^{(0)} \equiv 0$. The sequence (7) will be referred to as the *outer iteration loop* or the *nonlinear iteration loop*. Then, for every j the corresponding linear equation (7) is also solved iteratively as

$$\Delta E^{(j,k)} + k_0^2 E^{(j,k)} = -k_0^2 \epsilon |E^{(j-1,k)}|^{2\sigma} E^{(j,k-1)}, \quad k = 1, 2, \dots, K, \quad (8)$$

where $E^{(j,0)} = E^{(j-1,K)}$. We will refer to sequence (8) as to the *inner iteration loop*. It is clear that finding $E^{(j,k)}$ for every k in (8) amounts to solving the standard constant-coefficient inhomogeneous Helmholtz equation:

$$\Delta E + k_0^2 E = \Phi(x, z), \quad (9)$$

with $E = E^{(j,k)}(x, z)$ and $\Phi(x, z) \equiv -k_0^2 \epsilon |E^{(j-1,k)}(x, z)|^{2\sigma} E^{(j,k-1)}(x, z)$.

Eq. (9) is solved repeatedly [for updated $\Phi(x, z)$] on the rectangular computational domain via the separation of variables. In so doing, the boundary conditions in either the longitudinal direction z or the transverse direction x are specified for the linear constant-coefficient equation (9), rather than for the original nonlinear equation (6), see Section 2.2. The reason is that the key role of the boundary conditions is to distinguish between the waves propagating in different directions, see Section 1.2, and this is done most naturally in the linear framework. Once the iterations converge, the resulting solution is assumed to inherit the desired wave radiation properties that are built into the methodology for solving Eq. (9). We also note that no rigorous convergence theory is available for the nested iterations (7) and (8). The inner loop (8) can formally be interpreted as a fixed point scheme, see [6], but altogether the convergence of this iteration can only be judged experimentally in the discrete framework.

2.2. Separation of variables and boundary conditions

The Sommerfeld radiation boundary conditions in the transverse direction x are set based on factorization of the one-dimensional second-order Helmholtz operator into the product of two first-order factors:

$$\partial_{xx} + k_0^2 \mathbf{I} = (\partial_x + ik_0 \mathbf{I})(\partial_x - ik_0 \mathbf{I}), \quad (10)$$

where $i = \sqrt{-1}$, \mathbf{I} is the operator identity, and $k_0 = \sqrt{k_0^2} > 0$. While the equation $(\partial_{xx} + k_0^2 \mathbf{I})E = 0$ admits two linearly independent solutions, e^{ik_0x} and e^{-ik_0x} , each of the first-order factors on the right-hand side of (10) selects only one solution from the foregoing pair. Therefore, the boundary conditions

$$E_x - ik_0 E|_{x=x_{\max}} = 0 \quad \text{and} \quad E_x + ik_0 E|_{x=-x_{\max}} = 0 \quad (11)$$

correspond to propagation of waves only upward or only downward, respectively, see Fig. 1. In other words, they guarantee that the upper artificial boundary $x = x_{\max}$ be completely transparent for the plane waves e^{ik_0x} traveling with normal incidence in the positive x direction, and the lower artificial boundary $x = -x_{\max}$ be completely transparent for the plane waves e^{-ik_0x} traveling with normal incidence in the negative x direction.

Let us now introduce the following eigenvalue problem on $[-x_{\max}, x_{\max}]$ for the transverse Laplacian $\Delta_{\perp} \equiv \partial_{xx}$, subject to the radiation boundary conditions (11):

$$\psi_{xx} = \lambda \psi, \quad \psi_x - ik_0 \psi|_{x=x_{\max}} = 0, \quad \psi_x + ik_0 \psi|_{x=-x_{\max}} = 0. \quad (12)$$

It is easy to show that the eigenfunctions $\psi = \psi^{(l)}(x)$ of (12) are given by $\psi = \cosh(\sqrt{\lambda}x)$ and $\psi = \sinh(\sqrt{\lambda}x)$, and that the eigenvalues $\lambda = \lambda^{(l)}$ can be obtained by solving the transcendental equation:

$$e^{2\sqrt{\lambda}x_{\max}} = \pm \frac{\sqrt{\lambda} + ik_0}{\sqrt{\lambda} - ik_0}, \quad (13)$$

where the plus and minus signs correspond to $\psi = \cosh(\sqrt{\lambda}x)$ and to $\psi = \sinh(\sqrt{\lambda}x)$, respectively.

Eq. (12) is not a classical Sturm–Liouville problem, because the operator ∂_{xx} subject to boundary conditions (11) is not self-adjoint. As such, one should not expect its eigenvalues to be real. Indeed,

Proposition 1. *Let λ be an eigenvalue of (12). Then, $\text{Im}(\lambda) > 0$.*

Proof. Let $\sqrt{\lambda} = a + ib$, where $a, b \in \mathbb{R}$. Taking the absolute value of each side of Eq. (13) yields

$$e^{4ax_{\max}} = \frac{a^2 + (k_0 + b)^2}{a^2 + (k_0 - b)^2}.$$

If $a = 0$, then the left-hand-side term above is equal to one, hence $b = 0$. Therefore, the eigenfunctions are equal to a constant, which has to be zero because of the boundary conditions (11). Therefore, $a \neq 0$.

If $a > 0$, then the left-hand-side term above is greater than one, hence $b > 0$. Similarly, if $a < 0$ then $b < 0$. Therefore, $\text{Im}(\lambda) = 2ab > 0$. \square

Furthermore, since the boundary-value problem (12) is not self-adjoint, its eigenfunctions $\{\psi^{(l)}(x)\}$ are, generally speaking, nonorthogonal. Therefore, one cannot separate the variables in Eq. (9) using standard Fourier expansion of its solution in terms of the eigenfunctions of (12). It, however, turns out that one can still build the expansion $E(x, z) = \sum_l \hat{E}_l(z) \psi^{(l)}(x)$ with the help of the following “real orthogonality” property of the eigenfunctions $\{\psi^{(l)}(x)\}$ (see, e.g., [18] for further detail) summarized in Propositions 2 and 3:

Proposition 2. Let $\psi^{(m)}$ and $\psi^{(n)}$ be the eigenvectors of (12) with corresponding eigenvalues $\lambda_m \neq \lambda_n$. Then,

$$\int_{-x_{\max}}^{x_{\max}} \psi^{(m)}(x)\psi^{(n)}(x) dx = 0. \tag{14}$$

Proof. Since $\psi^{(m)} = \lambda_m^{-1} \psi_{xx}^{(m)}$,

$$\begin{aligned} \int_{-x_{\max}}^{x_{\max}} \psi^{(m)}\psi^{(n)} dx &= \lambda_m^{-1} \int_{-x_{\max}}^{x_{\max}} \psi_{xx}^{(m)}\psi^{(n)} dx = \lambda_m^{-1} \left(\psi_x^{(m)}\psi^{(n)} \Big|_{-x_{\max}}^{x_{\max}} - \int_{-x_{\max}}^{x_{\max}} \psi_x^{(m)}\psi_x^{(n)} dx \right) \\ &= \lambda_m^{-1} \left(ik_0\psi^{(m)}\psi^{(n)} \Big|_{-x_{\max}}^{x_{\max}} - \int_{-x_{\max}}^{x_{\max}} \psi_x^{(m)}\psi_x^{(n)} dx \right). \end{aligned}$$

Similarly,

$$\int_{-x_{\max}}^{x_{\max}} \psi^{(m)}\psi^{(n)} dx = \lambda_n^{-1} \left(ik_0\psi^{(m)}\psi^{(n)} \Big|_{-x_{\max}}^{x_{\max}} - \int_{-x_{\max}}^{x_{\max}} \psi_x^{(m)}\psi_x^{(n)} dx \right).$$

Therefore, the result follows. \square

Proposition 3. Let ψ be an eigenfunction of (12). Then,

$$\int_{-x_{\max}}^{x_{\max}} \psi^2(x) dx \neq 0.$$

Proof. Let $\psi = \cosh(\sqrt{\lambda}x)$. Then, using (13),

$$\begin{aligned} \int_{-x_{\max}}^{x_{\max}} \cosh^2(\sqrt{\lambda}x) dx &= \frac{1}{2} \int_{-x_{\max}}^{x_{\max}} [1 + \cosh(2\sqrt{\lambda}x)] dx = \frac{1}{2} \left[x + \frac{\sinh(2\sqrt{\lambda}x)}{2\sqrt{\lambda}} \right]_{-x_{\max}}^{x_{\max}} = x_{\max} + \frac{\sinh(2\sqrt{\lambda}x_{\max})}{2\sqrt{\lambda}} \\ &= x_{\max} + \frac{\exp(2\sqrt{\lambda}x_{\max}) - \exp(-2\sqrt{\lambda}x_{\max})}{4\sqrt{\lambda}} = x_{\max} + \frac{\frac{\sqrt{\lambda}+ik_0}{\sqrt{\lambda}-ik_0} - \frac{\sqrt{\lambda}-ik_0}{\sqrt{\lambda}+ik_0}}{4\sqrt{\lambda}} = x_{\max} + \frac{4\frac{\sqrt{\lambda}ik_0}{\lambda+k_0^2}}{4\sqrt{\lambda}} \\ &= x_{\max} + \frac{ik_0}{\lambda+k_0^2} = \frac{x_{\max}(\text{Re}(\lambda) + k_0^2) + i(x_{\max}\text{Im}(\lambda) + k_0)}{\lambda+k_0^2}. \end{aligned}$$

Since $\text{Im}(\lambda) > 0$, see Proposition 1, the imaginary part of the numerator is nonzero, and the result follows. The proof for $\psi = \sinh(\sqrt{\lambda}x)$ is similar. \square

Note that as the eigenfunctions $\psi^{(m)}$ and $\psi^{(n)}$ are complex, expression (14) does not yield a genuine inner product. However, Propositions 2 and 3 still indicate that the eigenfunctions of (12) can be rescaled so that

$$\int_{-x_{\max}}^{x_{\max}} \psi^{(m)}(x)\psi^{(n)}(x) dx = \delta_{mn}.$$

Therefore, for all $z \in [0, z_{\max}]$, solution to the linear Helmholtz equation (9) can, in principle, be obtained by expanding E and Φ on $[-x_{\max}, x_{\max}]$ in terms of the rescaled eigenfunctions $\psi^{(l)}(x)$ of (12), i.e.,

$$E(x, z) = \sum_l \hat{E}_l(z)\psi^{(l)}(x), \quad \Phi(x, z) = \sum_l \hat{\Phi}_l(z)\psi^{(l)}(x), \tag{15}$$

where

$$\hat{E}_l(z) = \int_{-x_{\max}}^{x_{\max}} E(x, z)\psi^{(l)}(x) dx, \quad \hat{\Phi}_l(z) = \int_{-x_{\max}}^{x_{\max}} \Phi(x, z)\psi^{(l)}(x) dx. \tag{16}$$

Naturally, these expansions only make sense provided that the series (15) and (16) converge. The most comprehensive justification for convergence would be based on completeness of the system of eigenfunctions $\{\psi^{(l)}(x)\}$. We, however, do not know whether this system is complete, and a full treatment of this important issue is certainly beyond the scope of the present paper. Let us only note that the question of completeness for systems of nonorthogonal eigenfunctions that do enjoy the real orthogonality property (14) has been studied in the literature. For example, Ching et al. provide a physics-based argument toward completeness for a similar but not identical to ours non-self-adjoint setting that involves a potential well with a discontinuity at the boundary [19]. A rigorous mathematical approach to completeness for the systems of eigenfunctions of non-Hermitian operators has been developed by Agranovich et al. [18,20]. A key consideration there is that the non-Hermitian operators that arise in the context of diffraction of time-harmonic waves can be interpreted as small perturbations of self-adjoint operators. As far as we are concerned, however, in order to justify the approach of our current paper we may not even have to require completeness of the system of eigenfunctions $\{\psi^{(l)}(x)\}$; it is sufficient to require that the series (15) and (16) converge for the solutions of interest.

Therefore, for our practical purposes, we will hereafter be assuming the aforementioned convergence. Then, substituting representation (15) into Eq. (9) and taking the “real inner product” (14) with $\psi^{(l)}$ we obtain the following set of ordinary differential equations with respect to the unknown quantities $\hat{E}_l(z)$:

$$\frac{d^2 \hat{E}_l}{dz^2} + (k_0^2 + \lambda^{(l)}) \hat{E}_l = \hat{\Phi}_l. \quad (17)$$

Since Proposition 1 shows that $\text{Im}(\lambda^{(l)}) > 0$, we conclude from Eq. (17) that introduction of the Sommerfeld radiation boundary conditions at the lateral boundaries results in the addition of a positive linear “damping” in the z direction. This conclusion has an intuitive explanation, since as the beam propagates in the z direction, the transverse radiation at $\pm x_{\max}$ will obviously cause a power drain away from the interval $[-x_{\max}, x_{\max}]$. As shown in [8], introduction of linear damping has a regularizing effect on the NLH (which, e.g., has allowed us to solve the linearly damped NLH for initial conditions that led to singularity formation in the corresponding linearly damped NLS). However, unlike in [8], where linear damping was essentially motivated by the physical process of absorption of waves by the medium through which they propagate, the “damping” in (17) has nothing to do with actual physical absorption or with power losses. This “damping” in the z direction is rather a manifestation of the power radiation at $\pm x_{\max}$. Another important difference is that unlike the case of physical damping, the magnitude of “damping” in (17) is determined by the transverse eigenvalue $\lambda^{(l)}$, i.e., it changes with the mode number l .

To complete the current illustrative section, we yet have to describe a key component of the algorithm – nonlocal ABCs at the boundaries $z = 0$ and $z = z_{\max}$. Each Eq. (17) is to be considered on the interval $[0, z_{\max}]$ and is to be supplemented by the boundary conditions at the endpoints $z = 0$ and $z = z_{\max}$. Assuming that $\text{supp } \hat{\Phi}_l(z) \subseteq [0, z_{\max}]$, i.e., that if extended beyond the interval $[0, z_{\max}]$ the solution $\hat{E}_l(z)$ would be governed by the homogeneous counterpart of Eq. (17), we employ the same considerations as those that led to the Sommerfeld conditions (11) and obtain:

$$\frac{d\hat{E}_l}{dz} + i\sqrt{k_0^2 + \lambda^{(l)}} \hat{E}_l|_{z=0} = 0 \quad \text{and} \quad \frac{d\hat{E}_l}{dz} - i\sqrt{k_0^2 + \lambda^{(l)}} \hat{E}_l|_{z=z_{\max}} = 0. \quad (18)$$

The boundary conditions (18) guarantee that all outgoing waves will leave the domain $[0, z_{\max}]$ with no reflection. Indeed, the left-traveling waves $e^{-i\sqrt{k_0^2 + \lambda^{(l)}}z}$ will propagate freely through the endpoint $z = 0$, and the right-traveling waves $e^{i\sqrt{k_0^2 + \lambda^{(l)}}z}$ will propagate freely through the endpoint $z = z_{\max}$.

Having obtained the ABCs (18) that allow for reflectionless propagation of all the outgoing waves, we now need to take into account the given incoming wave as well, see formula (2) and Fig. 1. The incoming

field $E^{(\text{inc, left})}(x) = \sum_l \hat{E}_l^{(\text{inc, left})} \psi^{(l)}(x)$ that impinges on the interface $z = 0$ from the left gives rise to a solution of the homogeneous constant-coefficient linear Helmholtz equation of the form

$$E(x, z) = \sum_l \hat{E}_l^{(\text{inc, left})} e^{i\sqrt{k_0^2 + \lambda^{(l)}}z} \psi^{(l)}(x). \quad (19)$$

After the separation of variables, each individual component $\hat{E}_l = \hat{E}_l^{(\text{inc, left})} e^{i\sqrt{k_0^2 + \lambda^{(l)}}z}$ can be substituted into the first relation of (18). Since this component satisfies $\frac{d\hat{E}_l}{dz} = i\sqrt{k_0^2 + \lambda^{(l)}}\hat{E}_l$ for any z , the aforementioned substitution yields the two-way boundary condition at $z = 0$:

$$\frac{d\hat{E}_l}{dz} + i\sqrt{k_0^2 + \lambda^{(l)}}\hat{E}_l|_{z=0} = 2i\sqrt{k_0^2 + \lambda^{(l)}}\hat{E}_l^{(\text{inc, left})}. \quad (20a)$$

Symmetrically, if there is also incoming radiation at the right interface $z = z_{\max} : E^{(\text{inc, right})}(x) = \sum_l \hat{E}_l^{(\text{inc, right})} \psi^{(l)}(x)$, then a similar procedure leads to the two-way boundary condition at $z = z_{\max}$ as well:

$$\frac{d\hat{E}_l}{dz} - i\sqrt{k_0^2 + \lambda^{(l)}}\hat{E}_l|_{z=z_{\max}} = -2i\sqrt{k_0^2 + \lambda^{(l)}}\hat{E}_l^{(\text{inc, right})}. \quad (20b)$$

Note that while being able to correctly prescribe the given incoming wave(s), boundary conditions (20) still retain the full radiation capability of boundary conditions (18). Indeed, any solution of type (2) identically satisfies boundary condition (20a). A similar property will obviously hold for boundary condition (20b) as well. The capability of properly handling the waves propagating through a given interface in both directions has prompted us in [6] to call boundary conditions (20) *the two-way ABCs*. In contradistinction to that, relations of type (18) that only guarantee the radiation of waves in one particular direction, are often referred to in the literature as *the one-way Helmholtz equations* [5,6].

Let us also emphasize that as the two-way ABCs (20) are obtained in the transformed space individually for every mode l , they would become *nonlocal* if transformed back to the original space. Further discussion on nonlocality of the boundary conditions can be found in Section 5.

3. Algorithm – finite-difference formulation

We now describe the finite-difference implementation of the algorithm. The NLH (6) is approximated on the uniform two-dimensional Cartesian grid with mesh sizes $h_x = x_{\max}/M$ and $h_z = z_{\max}/N$, so that:

$$\begin{aligned} x_m &= m \cdot h_x, & m &= -M, \dots, 0, \dots, M, \\ z_n &= n \cdot h_z, & n &= 0, \dots, N. \end{aligned} \quad (21)$$

3.1. Fourth order scheme

The discrete implementation of the algorithm is carried out with the fourth order of accuracy. As we shall see, the construction of the algorithm and its analysis are more complex for our fourth-order discretization than they would have been for a second order discretization. Nevertheless, this was a price worth paying, since higher-order approximations offer the possibility to take fewer points per wavelength. In addition, higher order offers the capability to better resolve the small-scale phenomenon of backscattering at the background of the larger forward-propagating wave.

For our algorithm, we have chosen the standard central-difference fourth-order scheme on the grid (21):

$$\begin{aligned} & \frac{-E_{m-2,n} + 16E_{m-1,n} - 30E_{m,n} + 16E_{m+1,n} - E_{m+2,n}}{12h_x^2} + \frac{-E_{m,n-2} + 16E_{m,n-1} - 30E_{m,n} + 16E_{m,n+1} - E_{m,n+2}}{12h_z^2} \\ & + k_0^2 \left(1 + \epsilon |E_{m,n}|^{2\sigma}\right) E_{m,n} = 0. \end{aligned} \quad (22)$$

Scheme (22) is written on the stencil that extends five grid nodes wide in each coordinate direction. Therefore, scheme (22) can only be written for the interior nodes of the grid (21) that are at least two nodes away from the boundary. Alternatively, two ghost nodes can be added to grid (21) from each side in each direction. In either case, the finite-difference equations to be solved near the boundaries of the computational domain will differ from the interior equation (22), and these special near-boundary equations shall be interpreted as discrete boundary conditions for the scheme (22).

To keep all the notations straightforward, let us introduce the ghost nodes (x_m, z_n) for $m = \pm(M+1)$ and $\pm(M+2)$, and $n = -2, -1, N+1$, and $N+2$. From here on, we will assume that the special near-boundary treatment shall apply to these ghost nodes, while the finite-difference equations (22) can keep their form on the entire grid (21). It is important to emphasize that system (22) itself is a system of fourth order finite-difference equations, and therefore it requires more boundary conditions than the original second order differential equation does, even though the former approximates the latter. More precisely, we will see that scheme (22) will require two boundary conditions at each boundary, whereas the original differential equation requires only one, see Section 2.2.

As indicated in Section 2, the boundary conditions are to be set for the linear constant coefficient equation solved repeatedly on the inner loop of the nested iteration scheme. Therefore, for the nonlinear finite-difference system (22) we introduce an iterative solver fully analogous to the one outlined in Section 2.1 for the continuous case (cf. formula (8)):

$$\begin{aligned} & \frac{-E_{m-2,n}^{(j,k)} + 16E_{m-1,n}^{(j,k)} - 30E_{m,n}^{(j,k)} + 16E_{m+1,n}^{(j,k)} - E_{m+2,n}^{(j,k)}}{12h_x^2} + \frac{-E_{m,n-2}^{(j,k)} + 16E_{m,n-1}^{(j,k)} - 30E_{m,n}^{(j,k)} + 16E_{m,n+1}^{(j,k)} - E_{m,n+2}^{(j,k)}}{12h_z^2} \\ & + k_0^2 E_{m,n}^{(j,k)} = -k_0^2 \epsilon |E_{m,n}^{(j-1,K)}|^{2\sigma} E_{m,n}^{(j,k-1)}, \end{aligned} \quad (23)$$

where $j = 1, 2, \dots$; $k = 1, 2, \dots, K$; $E_{m,n}^{(0,0)} = 0$; and $E_{m,n}^{(r,0)} = E_{m,n}^{(j-1,K)}$. Next, denoting $E_{m,n} = E_{m,n}^{(j,k)}$ and $\Phi_{m,n} = -k_0^2 \epsilon |E_{m,n}^{(j-1,K)}|^{2\sigma} E_{m,n}^{(j,k-1)}$, we arrive at the fourth-order central-difference approximation to the linear constant coefficient Helmholtz equation (9):

$$\begin{aligned} & \frac{-E_{m-2,n} + 16E_{m-1,n} - 30E_{m,n} + 16E_{m+1,n} - E_{m+2,n}}{12h_x^2} + \frac{-E_{m,n-2} + 16E_{m,n-1} - 30E_{m,n} + 16E_{m,n+1} - E_{m,n+2}}{12h_z^2} \\ & + k_0^2 E_{m,n} = \Phi_{m,n}. \end{aligned} \quad (24)$$

System (24) is to be supplemented by the boundary conditions and solved repeatedly in the course of the iteration (23) for updated $\Phi_{m,n}$.

3.2. Transverse boundary conditions

To set the discrete Sommerfeld radiation boundary conditions at $m = \pm M$, we first need to identify the waves that propagate upward and downward in the corresponding discrete framework. To do that, let us consider the one-dimensional discrete homogeneous Helmholtz equation (for clarity, we suppress here the subscript n)

$$\frac{-E_{m-2} + 16E_{m-1} - 30E_m + 16E_{m+1} - E_{m+2}}{12h_x^2} + k_0^2 E_m = 0. \quad (25)$$

Eq. (25) has a four-dimensional fundamental set of solutions: $\{q_1^m, q_1^{-m}, q_2^m, q_2^{-m}\}$, where $q_1, 1/q_1, q_2,$ and $1/q_2$ are the four roots of the algebraic characteristic equation that corresponds to the discretization (25):

$$-1 + 16q + (12\alpha_x^2 - 30)q^2 + 16q^3 - q^4 = 0. \tag{26}$$

The parameter

$$\alpha_x = h_x k_0 \tag{27}$$

is a measure of how well the waves are resolved by the transverse grid. Therefore, in what follows we can assume that

$$0 < \alpha_x \ll 1.$$

To find the roots of Eq. (26), we rewrite it as

$$\begin{aligned} -1 + 16q + (12\alpha_x^2 - 30)q^2 + 16q^3 - q^4 &= -(q - q_1)(q - q_1^{-1})(q - q_2)(q - q_2^{-1}) \\ &= -(q^2 - d_1q + 1)(q^2 - d_2q + 1) \\ &= -1 + (d_1 + d_2)q - (2 + d_1 + d_2)q^2 + (d_1 + d_2)q^3 - q^4 \\ &= 0, \end{aligned} \tag{28}$$

where

$$d_1 = q_1 + q_1^{-1} \quad \text{and} \quad d_2 = q_2 + q_2^{-1}. \tag{29}$$

Therefore,

$$d_1 + d_2 = 16 \quad \text{and} \quad -2 - d_1d_2 = 12\alpha_x^2 - 30, \tag{30}$$

which implies

$$d_1 = 8 - 6\sqrt{1 + \alpha_x^2/3} \quad \text{and} \quad d_2 = 8 + 6\sqrt{1 + \alpha_x^2/3}, \tag{31}$$

and finally we have:

$$q_1 = \frac{d_1 + \sqrt{d_1^2 - 4}}{2}, \quad q_1^{-1} = \frac{d_1 - \sqrt{d_1^2 - 4}}{2}, \quad q_2 = \frac{d_2 - \sqrt{d_2^2 - 4}}{2}, \quad q_2^{-1} = \frac{d_2 + \sqrt{d_2^2 - 4}}{2}. \tag{32}$$

In [6] we have shown that when $\alpha_x \ll 1$ (i.e., when the wavenumber k_0 is well resolved on the transverse grid), the roots q_1 and q_1^{-1} have unit magnitudes, $|q_1| = |q_1^{-1}| = 1$, and the solutions q_1^m and q_1^{-m} approximate the genuine traveling waves e^{ik_0x} and e^{-ik_0x} , respectively, with the fourth order of accuracy:

$$q_1^m = e^{ik_0h_x m} + O(\alpha_x^4), \quad q_1^{-m} = e^{-ik_0h_x m} + O(\alpha_x^4).$$

Regarding the second pair of roots, q_2 and q_2^{-1} , it was shown in [6] that $|q_2| < 1$ and $|q_2^{-1}| > 1$. Accordingly, solutions q_2^m and q_2^{-m} are always evanescent. These solutions are numerical artifacts, as they do not exist in the continuous context, see Section 2.2, and only appear because the discretization is fourth order, whereas the original differential equation is second order. Still, the presence of a second evanescent pair of waves requires special treatment at the boundary and necessitates setting an additional pair of the boundary conditions.

Consistently with the idea of Section 2.2, to guarantee the reflectionless propagation of the discrete outgoing waves through the lateral artificial boundaries, we need to require that the boundary at $m = M$ be transparent for the waves q_1^m traveling upward, and the boundary at $m = -M$ be transparent for the waves q_1^{-m} traveling downward. This requirement constitutes the Sommerfeld radiation principle in the fourth order discrete framework, and it cannot be either altered or relaxed in any way. As, however,

concerns the second, evanescent, pair of waves, q_2^m and q_2^{-m} , more flexibility can be exercised toward their near-boundary treatment. Indeed, these waves are non-physical and as such, we only need to ensure stability of the resulting overall discretization once the corresponding second pair of the boundary conditions has been chosen.

The most natural way to treat the evanescent waves would be to require that the boundary $m = M$ be transparent for the waves q_2^m that decay upward, and the boundary $m = -M$ be transparent for the waves q_2^{-m} that decay downward. We have employed a similar strategy in [6] for constructing the nonlocal ABCs, see also Section 3.7. Then, the overall solution near and beyond the boundary $m = M$ has to be a linear combination of these and only these two solutions from the fundamental set:

$$E_m = c_1 q_1^m + c_2 q_2^m. \quad (33)$$

Eq. (33) enforced for $m \geq M - 1$ (which is a formal definition of what “near and beyond” the boundary $m = M$ means) translates into:

$$\text{Rank} \begin{bmatrix} E_{M-1} & E_M & E_{M+1} & E_{M+2} \\ 1 & q_1 & q_1^2 & q_1^3 \\ 1 & q_2 & q_2^2 & q_2^3 \end{bmatrix} = 2. \quad (34)$$

Recasting (34) via the appropriate determinants, we arrive at the following form of the discrete one-way Helmholtz equation:

$$E_{M+1} = (q_1 + q_2)E_M - q_1 q_2 E_{M-1}, \quad (35a)$$

$$E_{M+2} = [(q_1 + q_2)^2 - q_1 q_2]E_M - (q_1 + q_2)q_1 q_2 E_{M-1}, \quad (35b)$$

that expresses the values of the solution at the ghost nodes $M + 1$ and $M + 2$ through its values on the interior grid.

Similarly, requiring that for $m \leq -M + 1$:

$$E_m = c_{-1} q_1^{-m} + c_{-2} q_2^{-m}, \quad (36)$$

we have:

$$\text{Rank} \begin{bmatrix} E_{-M-2} & E_{-M-1} & E_{-M} & E_{-M+1} \\ 1 & q_1^{-1} & q_1^{-2} & q_1^{-3} \\ 1 & q_2^{-1} & q_2^{-2} & q_2^{-3} \end{bmatrix} = 2, \quad (37)$$

and eventually obtain:

$$E_{-M-1} = (q_1 + q_2)E_{-M} - q_1 q_2 E_{-M+1}, \quad (38a)$$

$$E_{-M-2} = [(q_1 + q_2)^2 - q_1 q_2]E_{-M} - (q_1 + q_2)q_1 q_2 E_{-M+1}. \quad (38b)$$

It will be convenient to rewrite the one-way discrete Helmholtz equations (35) and (38) using universal notations:

$$\begin{aligned} E_{M+1} &= C_1 E_M + C_2 E_{M-1}, & E_{-M-1} &= C_1 E_{-M} + C_2 E_{-M+1}, \\ E_{M+2} &= C_3 E_M + C_4 E_{M-1}, & E_{-M-2} &= C_3 E_{-M} + C_4 E_{-M+1}, \end{aligned} \quad (39)$$

where $M + 1$, $M + 2$, $-M - 1$, and $-M - 2$ are the ghost nodes, and

$$\begin{aligned} C_1 &= q_1 + q_2, & C_3 &= (q_1 + q_2)^2 - q_1 q_2, \\ C_2 &= -q_1 q_2, & C_4 &= -(q_1 + q_2)q_1 q_2. \end{aligned} \quad (40)$$

Besides (35), (38), or equivalently (39), (40), there are many alternative approaches to treating the evanescent waves q_2^m and q_2^{-m} near the boundaries $m = \pm M$.⁴ In fact, the component $c_2 q_2^m$ in formula (33) can be replaced with almost any linear combination of q_2^m and q_2^{-m} , and the same is true regarding the component $c_{-2} q_2^{-m}$ in formula (36), as long as the chosen linear combinations are linearly independent. In so doing, stability will be guaranteed, except in some special degenerate cases, see [21].

In Section 3.5, we will formulate and analyze a discrete eigenvalue problem analogous to (12). To facilitate that analysis, let us now choose an alternative treatment of the evanescent waves, or in other words, select the constants C_1 , C_2 , C_3 , and C_4 , so that boundary conditions (39) would reduce to a particularly convenient symmetric form in the limiting case $\alpha_x = 0$. While keeping the same radiation conditions for q_1^m at $m = M$ and for q_1^{-m} at $m = -M$, let us require altogether for $m \geq M - 1$ (cf. formula (33)):

$$E_m = c_1 q_1^m + c_2 (q_2^m + q_2^{2M+1} q_2^{-m}), \tag{41}$$

and for $m \leq -M + 1$ (cf. formula (36)):

$$E_m = c_{-1} q_1^{-m} + c_{-2} (q_2^{2M+1} q_2^m + q_2^{-m}). \tag{42}$$

Then, taking the common factor q_2^{M-1} outside of the brackets in formula (41), we, instead of (34), obtain:

$$\text{Rank} \begin{bmatrix} E_{M-1} & E_M & E_{M+1} & E_{M+2} \\ 1 & q_1 & q_1^2 & q_1^3 \\ 1 + q_2^3 & q_2 + q_2^2 & q_2^2 + q_2 & q_2^3 + 1 \end{bmatrix} = 2. \tag{43}$$

Similarly, instead of (37) we have:

$$\text{Rank} \begin{bmatrix} E_{-M-2} & E_{-M-1} & E_{-M} & E_{-M+1} \\ q_1^3 & q_1^2 & q_1 & 1 \\ q_2^3 + 1 & q_2^2 + q_2 & q_2 + q_2^2 & 1 + q_2^3 \end{bmatrix} = 2. \tag{44}$$

Recasting (43) and (44) via determinants as before, we obtain the alternative one-way discrete Helmholtz equations in the same general form (39), except that instead of (40), the coefficients become:

$$\begin{aligned} C_1 &= \frac{q_2 + q_2^2 - q_1^2(1 + q_2^3)}{q_2 + q_2^2 - q_1(1 + q_2^3)}, & C_3 &= \frac{(1 + q_2^3)(1 - q_1^3)}{q_2 + q_2^2 - q_1(1 + q_2^3)}, \\ C_2 &= -\frac{(q_2 + q_2^2)(q_1 - q_1^2)}{q_2 + q_2^2 - q_1(1 + q_2^3)}, & C_4 &= -\frac{q_1(1 + q_2^3) - q_1^3(q_2 + q_2^2)}{q_2 + q_2^2 - q_1(1 + q_2^3)}. \end{aligned} \tag{45}$$

We re-emphasize that both boundary conditions (39), (40) and (39), (45) share the same fundamental property – they render the reflectionless radiation of the physical waves q_1^m and q_1^{-m} . The only difference between these boundary conditions is in how they address the evanescent modes q_2^m and q_2^{-m} . Moreover, both (39), (40) and (39), (45) yield an overall stable discretization (see Section 3.3). As such, one should not expect any major differences in the numerical performance of the algorithm equipped with either type of the boundary conditions (which we have indeed confirmed numerically). However, whereas the boundary conditions (39), (40) are most intuitive, the boundary conditions (39), (45) will be shown to possess additional symmetries which become useful when proving properties of the transverse eigenvectors in Section 3.5.

⁴ We recall that the Sommerfeld radiation principle leaves no flexibility in treating the physical waves q_1^m and q_1^{-m} .

3.3. Stability of the discrete approximation

A necessary condition for numerical stability of the overall discrete approximation is linear independence of the boundary conditions that we have chosen.

Proposition 4. *Boundary conditions (39), (40) or (39), (45) are linearly independent.*

Proof. Linear independence of boundary conditions (39), (40) is apparent, as due to (33) and (36) the homogeneous system (25) supplemented by (35) and (38) may only have a trivial solution $E_m = 0$ for $m = -M, \dots, M$. Linear independence of boundary conditions (39), (45) is a little more subtle. Let $E_m = c'_1 q_1^m + c'_{-1} q_1^{-m} + c'_2 q_2^m + c'_{-2} q_2^{-m}$ be a solution to Eq.(25) for $m = -M, \dots, M$, which also satisfies (41) for $M - 1 \leq m \leq M + 2$ and satisfies (42) for $-M - 2 \leq m \leq -M + 1$. Then, we immediately obtain $c'_1 = 0$ because this component is eliminated by (42), and $c'_{-1} = 0$ because this component is eliminated by (41). For the other pair of waves we have $c'_2 = c_2 = c_{-2} q_2^{2M+1}$ and $c'_{-2} = c_2 q_2^{2M+1} = c_{-2}$, which is clearly not possible unless either $q_2 = 1$ or $2M + 1 = 0$. As none of the latter holds, we conclude that altogether $c'_1 = c'_{-1} = c'_2 = c'_{-2} = 0$, i.e., that Eq.(25) supplemented by boundary conditions(39), (45) may only have a trivial solution. \square

Sufficient conditions for stability are somewhat more delicate, see [21]. Since reproducing the complete analysis of [21] is outside the scope of this paper, we will only outline its key elements and then summarize the results relevant to our specific systems (25), (39), (40) and (25), (39), (45). Let us focus on the evanescent waves that are treated differently by (39), (40) and by (39), (45). Requiring that the only evanescent component propagating upward be q_2^m , see formula (33), implies that the “remaining part” of the two-dimensional linear subspace $\text{span}\{q_2^m, q_2^{-m}\}$ in the four-dimensional fundamental space of solutions has to vanish at the upper transverse boundary. Of course, there is no way of saying unambiguously what this remaining part is, because anything linearly independent with q_2^m will be appropriate. We therefore need to see what the boundary condition at the opposite endpoint prescribes, and from formula (36) we conclude that the foregoing remaining part of $\text{span}\{q_2^m, q_2^{-m}\}$ has to be taken as $c_{-2} q_2^{-m}$, which is exactly what's allowed at the lower transverse boundary. In other words, the one-way Helmholtz equation (35) can be interpreted, in particular, as a zero Dirichlet boundary condition for the component q_2^{-m} at $m = M$, i.e., $c_{-2} q_2^{-m}|_{m=M} = 0$. Assume now that we have committed an error of magnitude ε in this Dirichlet boundary condition, i.e., $c_{-2} q_2^{-m}|_{m=M} = \varepsilon$. This error will be carried inward (i.e., downward) by the wave $c_{-2} q_2^{-m} = \varepsilon q_2^{M-m}$, and by the time it reaches the opposite endpoint $m = -M$, it will have the magnitude $|\varepsilon q_2^{2M}| \ll \varepsilon$, because $|q_2| < 1$. In other words, the error once committed at an endpoint in the homogeneous problem may not increase throughout the domain.

Analogously, requiring that the evanescent contribution to the solution near and beyond the boundary $m = M$ be $q_2^m + q_2^{2M+1} q_2^{-m}$, see formula (41), means that the remaining part of $\text{span}\{q_2^m, q_2^{-m}\}$, again, has to vanish. In this case, the aforementioned remaining part should be identified with $c_{-2}(q_2^{2M+1} q_2^m + q_2^{-m})$ according to (42). The corresponding homogeneous Dirichlet boundary condition should obviously read $c_{-2}(q_2^{2M+1} q_2^m + q_2^{-m})|_{m=M} = 0$, see formula (42), and if we commit an error of the overall magnitude ε , then we have $c_{-2}(q_2^{2M+1} q_2^m + q_2^{-m})|_{m=M} = \varepsilon$, and consequently, $c_{-2} = \varepsilon q_2^M (q_2^{4M+1} + 1)^{-1}$. Thus, the component of the error that may grow downward, i.e., toward the interior of the computational domain, will be $c_{-2} q_2^{2M+1} q_2^m = \varepsilon q_2^{3M+1} (q_2^{4M+1} + 1)^{-1} q_2^m$ and clearly, $\forall m = -M, \dots, M : |c_{-2} q_2^{2M+1} q_2^m| < \varepsilon |q_2^M| \ll \varepsilon$.

To complete the stability analysis, one will need to apply similar considerations to the propagating modes q_1^m and q_1^{-m} , and also study the inhomogeneous problem and analyze how the perturbations of its right-hand side specified on the interval $m = -M, \dots, M$ may affect the solution. Altogether, based on this analysis (see [21]), we can conclude that either of the two systems (25), (39), (40) and (25), (39), (45) is weakly stable. The term “weakly” means here that the constant in the stability inequality will not be completely grid independent, as required by the classical definition, but may rather grow at a slow poly-

mial rate with the increase of the grid dimension (or equivalently, decrease of the grid size). The source of this growth is, in fact, the growth of the perturbations of the right-hand side; it is accounted for by the presence of the unit magnitude “propagating” roots q_1 and q_1^{-1} of the characteristic equation (26). Numerical results of Section 4 corroborate, however, that the foregoing slow growth does not present any substantial difficulties for implementation.

3.4. Discrete eigenvalue problem

Let us now formulate a discrete eigenvalue problem analogous to (12). For that purpose, we first introduce a $(2M + 1) \times (2M + 1)$ transverse discretization matrix A that would contain the fourth order central difference approximation of ∂_{xx} at the interior nodes, see (24) or (25), and would also incorporate the boundary conditions (39). By eliminating the ghost variables E_{-M-2} , E_{-M-1} , E_{M+1} , and E_{M+2} , we obtain:

$$A = \frac{1}{12h_x^2} \begin{bmatrix} -30 + C_5 & 16 + C_6 & -1 & 0 & 0 & 0 & \dots \\ 16 - C_1 & -30 - C_2 & 16 & -1 & 0 & 0 & \dots \\ -1 & 16 & -30 & 16 & -1 & 0 & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \dots & 0 & -1 & 16 & -30 & 16 & -1 \\ \dots & 0 & 0 & -1 & 16 & -30 - C_2 & 16 - C_1 \\ \dots & 0 & 0 & 0 & -1 & 16 + C_6 & -30 + C_5 \end{bmatrix}, \tag{46a}$$

where

$$C_5 = 16C_1 - C_3, \quad C_6 = 16C_2 - C_4. \tag{46b}$$

The discrete counterpart of the eigenvalue problem (12) can now be formulated as follows:

$$A\psi = \lambda\psi. \tag{47}$$

In order to be able to separate the variables in the finite-difference system (24) using the eigenvectors of A , we need to make sure that these eigenvectors form a basis. To do that, we will use the results of the forthcoming Propositions 5 and 6 to prove in Proposition 7 that the $(2M + 1) \times (2M + 1)$ matrix A of (46) has $(2M + 1)$ distinct eigenvalues. This implies that A has $(2M + 1)$ linearly-independent eigenvectors. Then we will prove in Proposition 10 that these eigenvectors can be rescaled so that they possess the real orthogonality property, that allows us to arrive at formula (53) for the inverse of the matrix of the eigenvectors; the latter is needed for the actual implementation of the separation of variables.

Let us recall that the roots of the characteristic equation (26), and consequently, the entries of the matrix A of (46), are functions of α_x . In particular, it is easy to see that in the fine grid/long waves limit $\alpha_x \rightarrow 0+$, we have $q_1 \rightarrow 1$. For $q_1 = 1$ formulae (45) imply that $C_1 = C_4 = 1$ and $C_2 = C_3 = 0$. As such, boundary conditions (39), (45) for $\alpha_x = 0$ reduce to

$$\begin{aligned} E_{M+1} &= E_M, & E_{-M-1} &= E_{-M}, \\ E_{M+2} &= E_{M-1}, & E_{-M-2} &= E_{-M+1}. \end{aligned} \tag{48}$$

Equalities (48) imply symmetry with respect to the points $\pm(M + 1/2)$, which can also be interpreted as a fourth order approximation of the homogeneous Neumann boundary conditions. Note that boundary conditions (39), (40) do not reduce to the form as simple as (48) for $\alpha_x = 0$. This was the primary reason for introducing the new version (39), (45), rather than using the more intuitive version (39), (40). Moreover, when definition (45) is employed, the matrix A of (46) acquires a particularly convenient form for $\alpha_x = 0$:

$$A(0) = \frac{1}{12h_x^2} \begin{bmatrix} -14 & 15 & -1 & 0 & 0 & 0 & \dots \\ 15 & -30 & 16 & -1 & 0 & 0 & \dots \\ -1 & 16 & -30 & 16 & -1 & 0 & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \dots & 0 & -1 & 16 & -30 & 16 & -1 \\ \dots & 0 & 0 & -1 & 16 & -30 & 15 \\ \dots & 0 & 0 & 0 & -1 & 15 & -14 \end{bmatrix}, \tag{49}$$

so that the corresponding eigenvalue problem (47) can be solved analytically. This, again, would not have been the case if we used definition (40) instead of (45).

Proposition 5. *The eigenvalues $\lambda^{(l)}(0)$ and eigenvectors $\psi^{(l)}(0) = [\psi_{-M}^{(l)}(0), \dots, \psi_0^{(l)}(0), \dots, \psi_M^{(l)}(0)]^T$ of the matrix $A(0)$ of (49) are given by*

$$\lambda^{(l)}(0) = -\frac{16}{3h_x^2} \sin^2\left(\frac{1}{2} \frac{l\pi}{2M+1}\right) + \frac{1}{3h_x^2} \sin^2\left(\frac{l\pi}{2M+1}\right), \tag{50a}$$

$$\psi_m^{(l)}(0) = b_l \cos\left[l\left(\frac{\pi m}{2M+1} + \frac{\pi}{2}\right)\right], \tag{50b}$$

where $l = 0, 1, \dots, 2M$; $b_0 = \frac{1}{\sqrt{2M+1}}$, and $b_l = \sqrt{\frac{2}{2M+1}}$ for $l > 0$.

The eigenvectors $\psi^{(l)}(0)$ of (50b) are orthonormal, and the eigenvalues $\lambda^{(l)}(0)$ of (50a) are distinct.

Proof. Orthonormality of the eigenvectors (50b): $\sum_{m=-M}^M \psi_m^{(l)}(0)\psi_m^{(k)}(0) = \delta_{lk}$, as well as specific expressions for the eigenvalues (50a) and the normalization constants b_l in (50b), can be obtained with an argument almost identical to the one employed in [6, Section 5], and we will not reproduce it here (also see [22] for the general perspective on discrete Fourier series). We only need to show that the eigenvalues $\lambda^{(l)}(0)$ of (50a) are distinct. This is easy to see because (50a) implies

$$\begin{aligned} \lambda^{(l)}(0) &= -\frac{4}{3h_x^2} \sin^2\left(\frac{1}{2} \frac{l\pi}{2M+1}\right) \left[4 - \cos^2\left(\frac{1}{2} \frac{l\pi}{2M+1}\right)\right] \\ &= -\frac{4}{3h_x^2} \sin^2\left(\frac{1}{2} \frac{l\pi}{2M+1}\right) \left[3 + \sin^2\left(\frac{1}{2} \frac{l\pi}{2M+1}\right)\right], \end{aligned}$$

and as the argument of the sine function is always on the interval $[0, \pi/2)$, the quantities $\lambda^{(l)}(0)$ are monotonically decreasing in l . \square

Proposition 6. *The eigenvalues $\lambda^{(l)} = \lambda^{(l)}(\alpha_x)$ of the matrix A of (46), with $\{C_j\}_{j=1}^4$ given by (45), are differentiable functions of α_x at $\alpha_x = 0$.*

Proof. According to Proposition 5, when $\alpha_x = 0$ the eigenvalues $\lambda^{(l)}(0)$ of $A(0)$ of (49) are distinct. These eigenvalues are roots of the characteristic polynomial of the matrix that can be written as a product of monomials because of the aforementioned distinctness: $P_A(\lambda) = a(\lambda - \lambda^{(0)}(0)) \dots (\lambda - \lambda^{(2M)}(0))$. Then, for any given $l = 0, \dots, 2M$ we obviously have: $\frac{\partial P_A}{\partial \lambda} \Big|_{\alpha_x=0, \lambda=\lambda^{(l)}(0)} \neq 0$; and the desired result follows from the implicit function theorem. \square

Proposition 7. *Let $\alpha_x \ll 1$. Then, the $(2M+1) \times (2M+1)$ matrix A of (46), with $\{C_j\}_{j=1}^4$ given by (45), has $2M+1$ distinct eigenvalues $\lambda^{(l)} = \lambda^{(l)}(\alpha_x)$.*

Proof. According to Proposition 6, we can write for $l = 0, 1, \dots, 2M$,

$$\lambda^{(l)}(\alpha_x) = \lambda^{(l)}(0) + \left. \frac{d\lambda^{(l)}}{d\alpha_x} \right|_{\alpha_x=0} \alpha_x + O(\alpha_x^2). \tag{51}$$

For the matrix $A(0)$ of (49), its eigenvalues $\lambda^{(l)}(0)$ are given by (50a) and are distinct. Then, according to (51), $\lambda^{(l)}(\alpha_x)$ are distinct for small α_x as well. \square

Proposition 7 clearly implies that the matrix A of (46), (45) has $2M + 1$ linearly independent eigenvectors. Our next goal is to show that this basis of eigenvectors has the real orthogonality property (Proposition 10).

Proposition 8. *The matrix A of (46), where $\{C_j\}_{j=1}^4$ are given either by (40) or by (45), is symmetric, i.e., $A^T = A$.*

Proof. It is sufficient to show that $-C_1 = C_6$ or equivalently, $-C_1 = 16C_2 - C_4$. When $\{C_j\}_{j=1}^4$ are given by (40), what we need to prove is $q_1 + q_2 = -(q_1 + q_2)q_1q_2 + 16q_1q_2$. Following (29) and (30), we can write:

$$16 = d_1 + d_2 = q_1 + q_1^{-1} + q_2 + q_2^{-1} = q_1 + q_2 + \frac{q_1 + q_2}{q_1q_2} = (q_1 + q_2) \left(1 + \frac{1}{q_1q_2} \right) = (q_1 + q_2) \frac{1 + q_1q_2}{q_1q_2}.$$

Consequently, we indeed have $16q_1q_2 = (q_1 + q_2)(1 + q_1q_2)$.

When $\{C_j\}_{j=1}^4$ are given by (45), we obtain

$$\begin{aligned} C_1 - C_4 + 16C_2 &= (q_2 + q_2^2)(1 - q_1^3 - 16(q_1 - q_1^2)) + (1 + q_2^3)(q_1 - q_1^2) \\ &= (1 - q_1)[(q_2 + q_2^2)(1 + q_1 + q_1^2 - 16q_1) + q_1(1 + q_2^3)] \\ &= (1 - q_1)[(q_2 + q_2^2)(1 - d_1q_1 + q_1^2 - (16 - d_1)q_1) + q_1(1 + q_2 + q_2^2 + q_2^3)] \\ &= (1 - q_1)q_1[-d_2q_2 - d_2q_2^2 + 1 + q_2 + q_2^2 + q_2^3] \\ &= (1 - q_1)q_1[(1 - d_2q_2 + q_2^2) + (q_2 - d_2q_2^2 + q_2^3)] = 0, \end{aligned}$$

where we have, again, used (30), and have also taken into account that according to (28): $1 - d_1q_1 + q_1^2 = 0$ and $1 - d_2q_2 + q_2^2 = 0$. \square

Proposition 9. *Let ϕ and ψ be eigenvectors of the matrix A of (46) with the corresponding eigenvalues λ_ϕ and λ_ψ . If $\lambda_\phi \neq \lambda_\psi$, then, $\phi^T \psi = 0$.*

Proof. Using the definition of eigenvectors and eigenvalues (47), and the symmetry of A established in Proposition 8, we have:

$$\lambda_\phi \phi^T \psi = (A\phi)^T \psi = \phi^T A^T \psi = \phi^T A \psi = \lambda_\psi \phi^T \psi.$$

Consequently, $(\lambda_\phi - \lambda_\psi)\phi^T \psi = 0$, which does imply the “discrete real orthogonality” $\phi^T \psi = 0$ as long as $\lambda_\phi \neq \lambda_\psi$. \square

Remark. Proposition 9 is the discrete analogue of Proposition 2.

Since the eigenvectors of A are, generally speaking, complex, the operation $\phi^T \psi$ is not a genuine inner product of the two vectors. Hence, it is not obvious whether any nontrivial eigenvector ψ would satisfy $\psi^T \psi \neq 0$. However, for $\alpha_x \ll 1$ a continuation argument can be employed similar to the one used when proving Propositions 6 and 7. Indeed, when $\alpha_x = 0$ the matrix $A(0)$ is real symmetric, with real eigenvalues and real orthonormal eigenvectors. For the latter, the foregoing operation $\phi^T \psi$ does provide a genuine dot

product. Then, by continuity, for any $l = 0, \dots, 2M$ and $\alpha_x \ll 1$: $\psi^{(l)\text{T}}(\alpha_x)\psi^{(l)}(\alpha_x) \equiv \sum_{m=-M}^M \psi_m^{(l)}(\alpha_x)\psi_m^{(l)}(\alpha_x) \neq 0$, because $\sum_{m=-M}^M \psi_m^{(l)}(0)\psi_m^{(l)}(0) = 1$. A juxtaposition of this argument with the result of Proposition 9 yields the following

Proposition 10. *Let $\alpha_x \ll 1$. Then, the eigenvectors $\psi^{(l)}$, $l = 0, \dots, 2M$, of the matrix A of (46), where $\{C_j\}_{j=1}^4$ are given by (45), can be normalized so that $(\tilde{\psi}^{(l)})^T \tilde{\psi}^{(m)} = \delta_{lm}$.*

Indeed, we only need to choose the normalization constants as $\mu^{(l)} = \psi^{(l)\text{T}}\psi^{(l)}$, and then normalize: $\tilde{\psi}^{(l)} := \psi^{(l)} / \sqrt{\mu^{(l)}}$.

Previously, with the Dirichlet boundary conditions set at the transverse boundaries, we have used the discrete Fourier transform to separate the variables in system (24). In the Sommerfeld case we rather need to use the transformation by means of the eigenvectors of A . This transformation will be a discrete analogue of (15) and (16). Moreover, as formula (53) will show, its inverse is numerically straightforward and inexpensive.

Let us introduce a square matrix of order $2M + 1$ that would have the original non-normalized eigenvectors $\psi^{(l)}$ as its columns:

$$\Psi = [\psi^{(0)}, \dots, \psi^{(2M)}]. \tag{52}$$

Therefore,

$$A\Psi = \Psi A, \quad A = \begin{bmatrix} \lambda^{(0)} & 0 & 0 & 0 \\ 0 & \lambda^{(1)} & 0 & 0 \\ & & \ddots & \\ 0 & \dots & 0 & \lambda^{(2M)} \end{bmatrix}.$$

From Proposition 10 it follows that $\tilde{\Psi}^T \tilde{\Psi} = I$ and

$$\Psi^T \Psi = \begin{bmatrix} \mu^{(0)} & 0 & 0 & 0 \\ 0 & \mu^{(1)} & 0 & 0 \\ & & \ddots & \\ 0 & \dots & 0 & \mu^{(2M)} \end{bmatrix}.$$

Then, one can easily see that

$$\Psi^{-1} = (\Psi^T \Psi)^{-1} \Psi^T = \begin{bmatrix} 1/\mu^{(0)} & 0 & 0 & 0 \\ 0 & 1/\mu^{(1)} & 0 & 0 \\ & & \ddots & \\ 0 & \dots & 0 & 1/\mu^{(2M)} \end{bmatrix} \Psi^T. \tag{53}$$

The easy formula (53) is used in the computations of Section 4 to invert the matrix of eigenvectors Ψ .

3.5. Separation of variables

We now use the transformation matrices Ψ of (52) and Ψ^{-1} of (53) to separate the variables in the discrete system (24).

Let us define the $2M + 1$ -dimensional vectors \mathcal{E}_n and Φ_n for $n = 0, \dots, N$:

$$\begin{aligned} \mathcal{E}_n &= [E_{-M,n}, \dots, E_{0,n}, \dots, E_{M,n}]^T, \\ \Phi_n &= [\Phi_{-M,n}, \dots, \Phi_{0,n}, \dots, \Phi_{M,n}]^T. \end{aligned}$$

Then, system (24) subject to the transverse radiation boundary conditions (39), (45) can be recast as follows:

$$A\mathcal{E}_n + \frac{-\mathcal{E}_{n-2} + 16\mathcal{E}_{n-1} - 30\mathcal{E}_n + 16\mathcal{E}_{n+1} - \mathcal{E}_{n+2}}{12h_z^2} + k_0^2\mathcal{E}_n = \Phi_n. \tag{54}$$

In analogy with (15) and (16), let $\mathcal{E}_n = \Psi \hat{\mathcal{E}}_n \iff \hat{\mathcal{E}}_n = \Psi^{-1}\mathcal{E}_n$ and $\Phi_n = \Psi \hat{\Phi}_n \iff \hat{\Phi}_n = \Psi^{-1}\Phi_n$. In so doing, system (54) transforms into:

$$A\hat{\mathcal{E}}_n + \frac{-\hat{\mathcal{E}}_{n-2} + 16\hat{\mathcal{E}}_{n-1} - 30\hat{\mathcal{E}}_n + 16\hat{\mathcal{E}}_{n+1} - \hat{\mathcal{E}}_{n+2}}{12h_z^2} + k_0^2\hat{\mathcal{E}}_n = \hat{\Phi}_n. \tag{55}$$

The variables in (55) have separated, as the diagonal structure of the matrix A enables a natural decomposition of system (55) into a set of $2M + 1$ separate one-dimensional systems that would govern individual components of $\hat{\mathcal{E}}_n = [\hat{E}_{0,n}, \dots, \hat{E}_{2M,n}]^T$:

$$\frac{-\hat{E}_{l,n-2} + 16\hat{E}_{l,n-1} - 30\hat{E}_{l,n} + 16\hat{E}_{l,n+1} - \hat{E}_{l,n+2}}{12h_z^2} + (k_0^2 + \lambda^{(l)})\hat{E}_{l,n} = \hat{\Phi}_{l,n}, \quad l = 0, \dots, 2M, \tag{56}$$

here $\hat{\Phi}_{l,n}$, $l = 0, \dots, 2M$, are components of the vector $\hat{\Phi}_n$.

Each system (56) needs to be solved independently of the others, after which the solution in the original space can be reconstructed as $\mathcal{E}_n = \Psi \hat{\mathcal{E}}_n$ for every $n = 0, \dots, N$. For any $l = 0, \dots, 2M$, the corresponding system (56) is still subdefinite unless supplemented by the boundary conditions at $n = 0$ and $n = N$. We describe these discrete two-way ABCs in Section 3.7.

3.6. Perturbation analysis

In Section 2, we have shown that the introduction of the Sommerfeld transverse radiation boundary conditions can be interpreted as linear “damping” in the z direction. We now derive an equivalent result at the discrete level.

We first prove the following auxiliary statement:

Proposition 11. *Let $\lambda = \lambda(\alpha_x)$ be an eigenvalue of the matrix A of (46), where $\{C_j\}_{j=1}^4$ are given by either (40) or by (45). Then,*

$$\left. \frac{d\lambda}{d\alpha_x} \right|_{\alpha_x=0} = iv,$$

where $v \in \mathbb{R}$.

Proof. Since $q_{1,2} = q_{1,2}(\alpha_x)$ are given by (31) and (32), which, in particular, means that $q_1(\alpha_x) = 1 + i\alpha_x + O(\alpha_x^2)$ for $\alpha_x \ll 1$, we can see that

$$q_1(0) = 1, \quad \left. \frac{dq_1}{d\alpha_x} \right|_{\alpha_x=0} = i, \quad q_2(0) = 7 - \sqrt{48} \approx 0.07, \quad \left. \frac{dq_2}{d\alpha_x} \right|_{\alpha_x=0} = 0. \tag{57}$$

Therefore, for either (40) or (45) it follows that

$$\left. \frac{dC_j}{d\alpha_x} \right|_{\alpha_x=0} = \left. \frac{\partial C_j}{\partial q_1} \frac{dq_1}{d\alpha_x} \right|_{\alpha_x=0} = i \left. \frac{\partial C_j}{\partial q_1} \right|_{q_1=1, q_2=7-\sqrt{48}}. \tag{58}$$

We thus see that $\frac{dC_j}{d\alpha_x}|_{\alpha_x=0}$, where $j = 1, \dots, 6$, is purely imaginary.

Let $\psi = \psi(\alpha_x)$ be the eigenvector corresponding to $\lambda(\alpha_x)$, rescaled such that $\psi^T \psi = 1$. Therefore,

$$0 = \frac{d}{d\alpha_x} \psi^T \psi = 2\psi^T \frac{d\psi}{d\alpha_x} \tag{59}$$

and

$$\lambda = \psi^T A \psi.$$

Differentiating the last relation with respect to α_x , using Eq. (59) and the symmetry of A (Proposition 8) we obtain

$$\frac{d\lambda}{d\alpha_x} = \psi^T \frac{dA}{d\alpha_x} \psi + \frac{d\psi^T}{d\alpha_x} A \psi + \psi^T A \frac{d\psi}{d\alpha_x} = \psi^T \frac{dA}{d\alpha_x} \psi + 2 \frac{d\psi^T}{d\alpha_x} A \psi = \psi^T \frac{dA}{d\alpha_x} \psi + 2\lambda \frac{d\psi^T}{d\alpha_x} \psi = \psi^T \frac{dA}{d\alpha_x} \psi. \tag{60}$$

Next, by (46) and the symmetry of A ,

$$\frac{dA}{d\alpha_x} = \frac{1}{12h_x^2} \begin{bmatrix} \frac{dC_5}{d\alpha_x} & -\frac{dC_1}{d\alpha_x} & 0 & 0 & 0 & 0 & \dots \\ -\frac{dC_1}{d\alpha_x} & -\frac{dC_2}{d\alpha_x} & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \dots & 0 & 0 & 0 & 0 & 0 & 0 \\ \dots & 0 & 0 & 0 & 0 & -\frac{dC_2}{d\alpha_x} & -\frac{dC_1}{d\alpha_x} \\ \dots & 0 & 0 & 0 & 0 & -\frac{dC_1}{d\alpha_x} & \frac{dC_5}{d\alpha_x} \end{bmatrix}. \tag{61}$$

When $\alpha_x = 0$, the matrix $A(0)$ is self-adjoint, hence its rescaled eigenvector $\psi(0)$ is real (see Proposition 5). Therefore, it follows from (58), (60) and (61) that $\frac{d\lambda}{d\alpha_x}(0)$ is purely imaginary. \square

Proposition 11 shows that to leading order, the small nonorthogonality parameter α_x affects only the imaginary part of the eigenvalue.⁵ We now show that as in the continuous case (cf. Proposition 1), the imaginary part is positive:

Proposition 12. $\nu > 0$.

Proof. Using (31), (32), (45), (46b) and (57) we can carry out explicitly the differentiation in (61), which yields

$$\frac{dA}{d\alpha_x} \Big|_{\alpha_x=0} \approx \frac{i}{12h_x^2} \begin{bmatrix} 14.08 & -1.08 & 0 & 0 & 0 & 0 & \dots \\ -1.08 & 0.083 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \dots & 0 & 0 & 0 & 0 & 0 & 0 \\ \dots & 0 & 0 & 0 & 0 & 0.083 & -1.08 \\ \dots & 0 & 0 & 0 & 0 & -1.08 & 14.08 \end{bmatrix}.$$

⁵ The effect on the real part of the eigenvalue is $O(\alpha_x^2)$.

Therefore, from (60),

$$\text{iv}^{(l)} = \frac{d\lambda^{(l)}}{d\alpha_x}(0) \approx \frac{i}{12h_x^2} \left[14\left(\psi_{-M}^{(l)}(0)\right)^2 + 14\left(\psi_M^{(l)}(0)\right)^2 - 2\psi_{-M}^{(l)}(0)\psi_{-M+1}^{(l)}(0) - 2\psi_M^{(l)}(0)\psi_{M-1}^{(l)}(0) \right].$$

Using (50b),

$$\frac{\psi_{\pm(M-1)}^{(l)}}{\psi_{\pm M}^{(l)}} = \frac{\cos(3\theta)}{\cos\theta}, \quad \theta = \frac{\pi}{2} \frac{l}{2M+1},$$

where $0 \leq \theta < \pi/2$. Since

$$14\cos^2\theta - 2\cos\theta\cos(3\theta) = 14\cos^2\theta - 2\cos(4\cos^3\theta - 3\cos\theta) = 20\cos^2\theta - 8\cos^4\theta > 0,$$

the result follows. \square

3.7. Nonlocal ABCs

The following construction of the ABCs at $n = 0$ and $n = N$ will, in fact, be very similar to the construction of the foregoing transverse radiation boundary conditions, see Section 3.2, and will basically be identical to the construction of the two-way ABCs in [6]. The only difference between the forthcoming derivation and its counterpart in [6] is of technical nature. Namely, in the current paper we are using the ghost nodes, see Section 3.1, which we did not do in [6]. We think that the use of the ghost nodes simplifies the presentation. As in Section 2.2, the role of the ABCs is to guarantee the reflectionless propagation of the discrete outgoing waves through the boundaries $n = 0$ and $n = N$, and at the same time, to prescribe the given incoming radiation.

Having separated the variables in system (24), we obtain a set of one-dimensional finite-difference equations (56). Each of these equations will be supplemented by 1D boundary conditions at $n = 0$ and at $n = N$ in much the same way as done in Section 3.2 for Eq. (25). Thus, let $q_1, 1/q_1, q_2$ and $1/q_2$ be the four roots of the algebraic characteristic equation that corresponds to the homogeneous counterpart of equation l of (56)⁶ (cf. Eq. (26)):

$$-1 + 16q + (12\alpha_z^2 - 30)q^2 + 16q^3 - q^4 = 0, \tag{62}$$

where (cf. formula (27))

$$\alpha_z = h_z \sqrt{k_0^2 + \lambda^{(l)}}. \tag{63}$$

For the homogeneous counterpart of this equation l of (56) we build the discrete one-way Helmholtz equation at the right endpoint $n = N$:

$$\hat{E}_{l,N+1} = (q_1 + q_2)\hat{E}_{l,N} - q_1q_2\hat{E}_{l,N-1}, \tag{64a}$$

$$\hat{E}_{l,N+2} = [(q_1 + q_2)^2 - q_1q_2]\hat{E}_{l,N} - (q_1 + q_2)q_1q_2\hat{E}_{l,N-1}, \tag{64b}$$

and at the left endpoint $n = 0$:

$$\hat{E}_{l,-1} = (q_1 + q_2)\hat{E}_{l,0} - q_1q_2\hat{E}_{l,1}, \tag{65a}$$

$$\hat{E}_{l,-2} = [(q_1 + q_2)^2 - q_1q_2]\hat{E}_{l,0} - (q_1 + q_2)q_1q_2\hat{E}_{l,1}. \tag{65b}$$

⁶ Note that unlike in Section 3.2, the parameter α_z of (63) and the roots of Eq. (62) all depend on the transverse mode number l .

Formulae (64) are analogous to (35), and formulae (65) are analogous to (38). In the capacity of boundary conditions, Eqs. (64) will facilitate reflectionless propagation of all the waves traveling in the positive z direction through the boundary $z = z_{\max}$, i.e., $n = N$, and Eqs. (65) will facilitate reflectionless propagation of all the waves traveling in the negative z direction through the boundary $z = 0$, i.e., $n = 0$.

By construction, the one-way discrete Helmholtz equations (65) and (64) can only be used as radiation boundary conditions for the homogeneous discrete Helmholtz equation. However, Eq. (56) are inhomogeneous. To take into account the inhomogeneity $\hat{\Phi}_{l,n}$ on the right-hand side of (56), we employ the discrete Green's function G^n , which is defined as solution to

$$\frac{-G^{n-2} + 16G^{n-1} - 30G^n + 16G^{n+1} - G^{n+2}}{12h_z^2} + (k_0^2 + \lambda^{(l)})G^n = \delta_n, \quad (66)$$

$$\delta_n = \begin{cases} 1, & n = 0, \\ 0, & n \neq 0, \end{cases}$$

satisfying the radiation principle, i.e., such that (cf. formulae (33) and (36)):

$$G^n = \begin{cases} a_1 q_1^n + a_2 q_2^n, & n \geq 0, \\ b_1 q_1^{-n} + b_2 q_2^{-n}, & n \leq 0. \end{cases} \quad (67)$$

The Green's function G^n of (66), (67) can be obtained in the closed form, the corresponding expressions for the coefficients a_1 , a_2 , b_1 , and b_2 can be found in [6]. This Green's function is defined for the entire range $-\infty < n < \infty$, and is, in fact, symmetric, i.e., $G^{-n} = G^n$ (or equivalently, $a_1 = a_2$ and $b_1 = b_2$).

With the help of the Green's function G^n of (66), (67), one can write down solution of equation l from (56) in the form of the discrete convolution:

$$\hat{E}_{l,n} = \sum_{j=0}^{j=N} \hat{\Phi}_{l,j} G^{n-j}. \quad (68)$$

The grid function $\hat{E}_{l,n}$ of (68) is also defined for $-\infty < n < \infty$. It satisfies the inhomogeneous equation l of (56) on the interval $0 \leq n \leq N$; and on the intervals of homogeneity it obviously satisfies (again, cf. (33) and (36)):

$$\hat{E}_{l,n} = \begin{cases} c_1 q_1^n + c_2 q_2^n, & n \geq N, \\ c_{-1} q_1^{-n} + c_{-2} q_2^{-n}, & n \leq 0. \end{cases}$$

Next, substituting expression (68) into (64a) we obtain:

$$\begin{aligned} \hat{E}_{l,N+1} - (q_1 + q_2)\hat{E}_{l,N} + q_1 q_2 \hat{E}_{l,N-1} &= \sum_{j=0}^{j=N} \hat{\Phi}_{l,j} G^{N+1-j} - (q_1 + q_2) \sum_{j=0}^{j=N} \hat{\Phi}_{l,j} G^{N-j} + q_1 q_2 \sum_{j=0}^{j=N} \hat{\Phi}_{l,j} G^{N-1-j} \\ &= \sum_{j=0}^{j=N} \hat{\Phi}_{l,j} [G^{N+1-j} - (q_1 + q_2)G^{N-j} + q_1 q_2 G^{N-1-j}] \\ &= \hat{\Phi}_{l,N} [G^1 - (q_1 + q_2)G^0 + q_1 q_2 G^{-1}], \end{aligned} \quad (69a)$$

and similarly, substituting it into (64b) we get:

$$\begin{aligned}
& \hat{E}_{l,N+2} - [(q_1 + q_2)^2 - q_1 q_2] \hat{E}_{l,N} + (q_1 + q_2) q_1 q_2 \hat{E}_{l,N-1} \\
&= \sum_{j=0}^{j=N} \hat{\Phi}_{l,j} \left[G^{N+2-j} - [(q_1 + q_2)^2 - q_1 q_2] G^{N-j} + (q_1 + q_2) q_1 q_2 G^{N-1-j} \right] \\
&= \hat{\Phi}_{l,N} \left[G^2 - [(q_1 + q_2)^2 - q_1 q_2] G^0 + (q_1 + q_2) q_1 q_2 G^{-1} \right].
\end{aligned} \tag{69b}$$

We note that in formulae (69a) and (69b) all terms except $j = N$ turn into zero because for all $j < N$ the shifted Green's function G^{N-j} clearly satisfies the homogeneous relations (64a) and (64b).

An identical argument applied near the left endpoint $n = 0$ yields instead of (65a) and (65b):

$$\hat{E}_{l,-1} - (q_1 + q_2) \hat{E}_{l,0} + q_1 q_2 \hat{E}_{l,1} = \hat{\Phi}_{l,0} [G^{-1} - (q_1 + q_2) G^0 + q_1 q_2 G^1] \tag{70a}$$

and

$$\hat{E}_{l,-2} - [(q_1 + q_2)^2 - q_1 q_2] \hat{E}_{l,0} + (q_1 + q_2) q_1 q_2 \hat{E}_{l,1} = \hat{\Phi}_{l,0} \left[G^{-2} - [(q_1 + q_2)^2 - q_1 q_2] G^0 + (q_1 + q_2) q_1 q_2 G^1 \right]. \tag{70b}$$

It only remains now to take into account the incoming waves. This is done similarly to the continuous case described in Section 2.2. Using the notations of Section 3.5, we transform the discrete field impinging from the left: $\hat{\mathcal{E}}^{(\text{inc, left})} = \Psi^{-1} \mathcal{E}^{(\text{inc, left})}$, and substitute the resulting individual modes $\hat{E}_l^{(\text{inc, left})} q_1^n$ into the left-hand side of the corresponding relations (70), which yields:

$$\hat{E}_{l,-1} - (q_1 + q_2) \hat{E}_{l,0} + q_1 q_2 \hat{E}_{l,1} = \hat{\Phi}_{l,0} [G^{-1} - (q_1 + q_2) G^0 + q_1 q_2 G^1] + \hat{E}_l^{(\text{inc, left})} [q_1^{-1} - (q_1 + q_2) + q_1^2 q_2] \tag{71a}$$

and

$$\begin{aligned}
\hat{E}_{l,-2} - [(q_1 + q_2)^2 - q_1 q_2] \hat{E}_{l,0} + (q_1 + q_2) q_1 q_2 \hat{E}_{l,1} &= \hat{\Phi}_{l,0} \left[G^{-2} - [(q_1 + q_2)^2 - q_1 q_2] G^0 + (q_1 + q_2) q_1 q_2 G^1 \right] \\
&+ \hat{E}_l^{(\text{inc, left})} \left[q_1^{-2} - [(q_1 + q_2)^2 - q_1 q_2] + (q_1 + q_2) q_1^2 q_2 \right].
\end{aligned} \tag{71b}$$

Similarly, at the right endpoint $n = N$ we substitute $\hat{E}_l^{(\text{inc, right})} q_1^{-n}$ and obtain instead of (69):

$$\begin{aligned}
\hat{E}_{l,N+1} - (q_1 + q_2) \hat{E}_{l,N} + q_1 q_2 \hat{E}_{l,N-1} &= \hat{\Phi}_{l,N} [G^1 - (q_1 + q_2) G^0 + q_1 q_2 G^{-1}] \\
&+ \hat{E}_l^{(\text{inc, right})} q_1^{-N} [q_1^{-1} - (q_1 + q_2) + q_1^2 q_2]
\end{aligned} \tag{72a}$$

and

$$\begin{aligned}
\hat{E}_{l,N+2} - [(q_1 + q_2)^2 - q_1 q_2] \hat{E}_{l,N} + (q_1 + q_2) q_1 q_2 \hat{E}_{l,N-1} \\
= \hat{\Phi}_{l,N} \left[G^2 - [(q_1 + q_2)^2 - q_1 q_2] G^0 + (q_1 + q_2) q_1 q_2 G^{-1} \right] \\
+ \hat{E}_l^{(\text{inc, right})} q_1^{-N} [q_1^{-2} - [(q_1 + q_2)^2 - q_1 q_2] + (q_1 + q_2) q_1^2 q_2].
\end{aligned} \tag{72b}$$

Inhomogeneous relations (71) and (72) provide the desired finite-difference two-way ABCs at the left and right boundaries, respectively. Altogether, to obtain the solution on every iteration, we now need to solve Eq. (56) for $n = 0, \dots, N$ and for all l , subject to the boundary conditions (71) that supplement the ghost nodes $n = -2$ and $n = -1$, and boundary conditions (72) that supplement the ghost nodes $n = N + 1$ and $n = N + 2$.

The ABCs (71) and (72) are fourth order discrete counterparts of the continuous boundary conditions (20). We emphasize that the right-hand sides in relations (71) and (72) are obtained by local formulae and as such, are numerically inexpensive. We also emphasize that similarly to boundary conditions (20), the discrete two-way ABCs (71) and (72) are obtained in the transformed space independently for each one-dimensional system (56), and would therefore become nonlocal if transformed back to the original space by means of the matrix Ψ . However, we never actually need to explicitly transform the ABCs to the original physical space, because system (24) is solved on each iteration (23) using the separation of variables (55) and consequently, the boundary conditions only need to be applied to variables $\hat{E}_{l,n}$ in the transformed space.

3.8. Resonances

If $\alpha_z = 0$, two roots q_1 and $1/q_1$ of the characteristic equation (62) merge into one double root $q_1 = 1/q_1 = 1$. Then, the homogeneous one-way discrete Helmholtz equations (65) and (64) become linearly dependent, which basically ruins the stability, see Section 3.3. For the previously used Dirichlet boundary conditions, all eigenvalues $\lambda^{(l)}$ were real and negative. Therefore, according to formula (63), having $\alpha_z = 0$ was indeed a possibility. An exact equality, of course, could always be avoided by slightly changing the parameters of the discretization in the transverse direction x , see [6]. However, in our numerical simulations we have observed a number of cases when instability manifested itself even for the modes with α_z close, rather than exactly equal, to zero. Some of those cases are reported in Section 4.2 and are referred to as resonances. On the other hand, according to Proposition 12, the Sommerfeld boundary conditions guarantee that all eigenvalues $\lambda^{(l)}$ have a non-trivial imaginary part.⁷ This allows one to separate α_z from zero for all l , see formula (63). and indeed, no instabilities of the aforementioned type have ever been seen in our numerical experiments with the Sommerfeld boundary conditions, see Section 4.2.

4. Results of computations

In this section, we present computational results obtained with the help of the algorithm of Section 3. In order to judge whether the iterations (23) converge, we monitor the following three quantities on the outer loop:

- (1) The difference between successive iterations: $\max_{m,n} |E_{m,n}^{(j)} - E_{m,n}^{(j-1)}|$.
- (2) The maximal residual: $\max_{m,n} |R_{m,n}^{(j)}|$, where

$$R_{m,n}^{(j)} = \frac{-E_{m-2,n}^{(j)} + 16E_{m-1,n}^{(j)} - 30E_{m,n}^{(j)} + 16E_{m+1,n}^{(j)} - E_{m+2,n}^{(j)}}{12h_x^2} + \frac{-E_{m,n-2}^{(j)} + 16E_{m,n-1}^{(j)} - 30E_{m,n}^{(j)} + 16E_{m,n+1}^{(j)} - E_{m,n+2}^{(j)}}{12h_z^2} + k_0^2 \left[1 + \epsilon |E_{m,n}^{(j)}|^{2\sigma} \right] E_{m,n}^{(j)}.$$

- (3) The mean residual: $\frac{1}{MN} \sum_m \sum_n |R_{m,n}^{(j)}|$.

⁷ Moreover, since $\text{Re}(\alpha_z) \approx 0$ implies that $\text{Re}(\lambda_l) \approx -k_0^2$, this case corresponds to modes propagating nearly-perpendicular to the z direction. Therefore, linear “damping” would be the largest for these modes.

4.1. Critical case

We solve the 2D NLH (6) in the critical case $\sigma = 2$ with $k_0 = 8$ and with the incident Gaussian input beam given by

$$E^{(\text{inc, left})}(x, 0) = e^{-x^2}.$$

As shown in [6], the fractional critical power of this input beam is equal to

$$p = \sqrt{\frac{2\epsilon}{3\pi}} k_0.$$

The highest input power that we could solve for in [6] was $\epsilon = 0.06$, which implies $p = 90\%$, i.e., the power of the corresponding NLS solution is 90% of the threshold power for collapse N_c .

We now demonstrate that when using the Sommerfeld boundary conditions at the transverse boundary one obtains a considerably more robust numerical algorithm, compared with the original algorithm with Dirichlet boundary conditions of [6]. The improved robustness manifests itself in two ways:

- (1) Better quality of the solution when both algorithms use the same numerical parameters – domain size and grid resolution.
- (2) Better convergence of iterations (23), i.e., a smaller transverse size of the computational domain is sufficient for the iterations to converge.

Our starting point for comparison of the two algorithms are the results of computations performed in [6]. In particular, we have determined in [6] that if the focusing beam in the critical regime is coming from the left interface $z = 0$, then the performance of the nonlocal ABCs at the opposite “downstream” boundary z_{\max} improves when this boundary is located not too close to the focusing zone; ideally in the region where the effect of nonlinearity becomes small. For $k_0 = 8$ and $\epsilon = 0.06$, the value of $z_{\max} = 40$ has been found sufficiently large, and we fix it for the computations in this paper as well. The transverse size of the computational domain x_{\max} has been shown to exert a critical influence on both the solution quality and the convergence of iterations in [6]; most likely because of the effect of wave reflections from the lateral boundaries equipped with the Dirichlet conditions. Below, we will carefully examine the influence of x_{\max} in the case of the Sommerfeld transverse boundary conditions. We have also conducted grid convergence studies in [6]. The solution has been shown to be more sensitive to the longitudinal wave resolution λ_0/h_z than to the transverse wave resolution λ_0/h_x , where the wavelength is given by $\lambda_0 = 2\pi/k_0$, the reason being that the beam dynamics primarily develops in the z direction. Starting with $h_z = \lambda_0/10$ and $h_x = \lambda_0/4$ the results become quite acceptable [6], while the values of $h_z = \lambda_0/20$ and $h_x = \lambda_0/8$ have been found sufficiently small so that further grid refinement did not provide any noticeable improvement of the solution quality.

We first solve the critical NLH with the Dirichlet transverse boundary conditions for $h_z = \lambda_0/20$ and $h_x = \lambda_0/8$. The minimum transverse size x_{\max} for which the iteration scheme (23) converged was found to be $x_{\max} = 13$; for smaller x_{\max} the iterations diverged (this divergence was found to be more rapid on finer grids, which is a natural phenomenon to expect). Convergence of the residuals was linear, see Fig. 2, with roughly forty outer-loop cycles required to obtain ten orders of magnitude reduction. All three convergence criteria show a similar convergence rate. The converged solution (magnitude of the electric field) for $x_{\max} = 13$ is shown in Figs. 3 and 4. Overall, the solution first undergoes a moderate self-focusing, and then starts to diffract. One can clearly see that reflections of the diffracting part of the beam from the lateral boundaries of the computational domain tend to corrupt the solution for larger z 's.

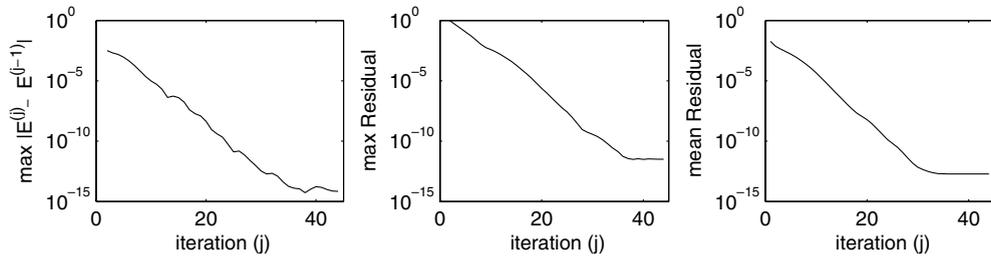


Fig. 2. Convergence history plots for the critical NLH subject to the Dirichlet transverse boundary conditions; $x_{\max} = 13$.

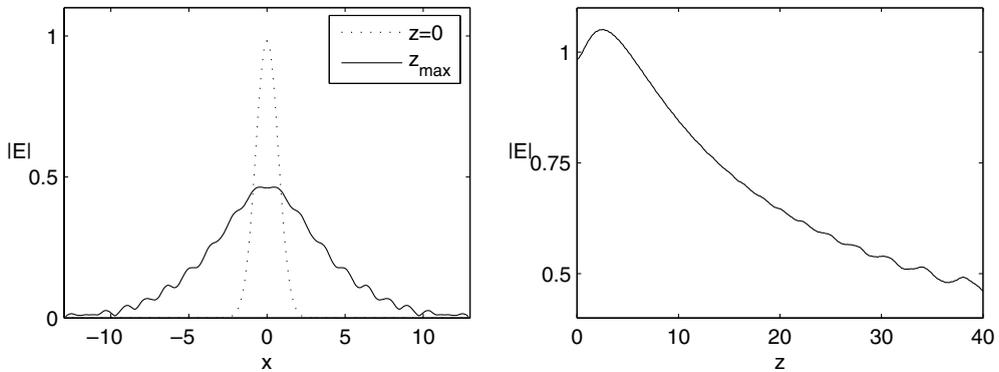


Fig. 3. Solution of the critical NLH subject to the Dirichlet transverse boundary conditions for $x_{\max} = 13$: Incoming and outgoing cross-sections (left) and on-axis amplitude $|E(0,z)|$ (right).

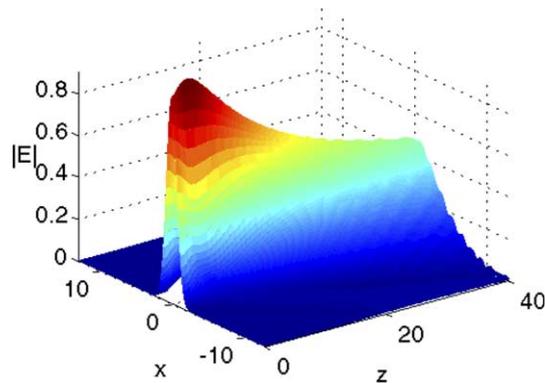


Fig. 4. Solution of the critical NLH subject to the Dirichlet transverse boundary conditions for $x_{\max} = 13$.

We now compare the foregoing results with those obtained with the Sommerfeld transverse boundary conditions, keeping all other parameters unchanged, see Figs. 6 and 7. While the iteration convergence rate remains roughly the same, see Fig. 5, the quality of the solution has obviously much improved, and the lateral reflections due to the Dirichlet boundary conditions are significantly reduced.

A grid convergence study for the new Sommerfeld boundary conditions is obviously needed to corroborate the validity of the results. We have recomputed the previous solution for $z_{\max} = 40$ and $x_{\max} = 13$ on a

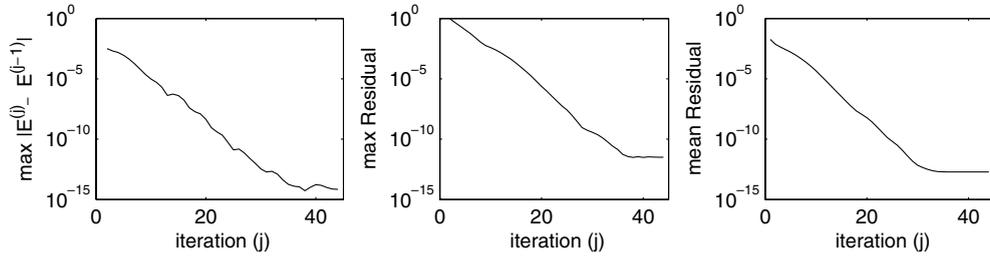


Fig. 5. Convergence history plots for the critical NLH subject to the Sommerfeld transverse boundary conditions; $x_{\max} = 13$.

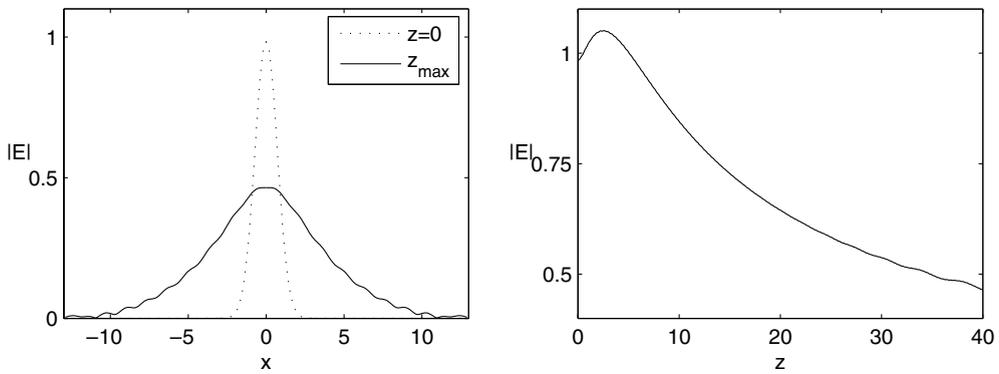


Fig. 6. Solution of the critical NLH subject to the Sommerfeld transverse boundary conditions for $x_{\max} = 13$: Incoming and outgoing cross-sections (left) and on-axis amplitude $|E(0,z)|$ (right).

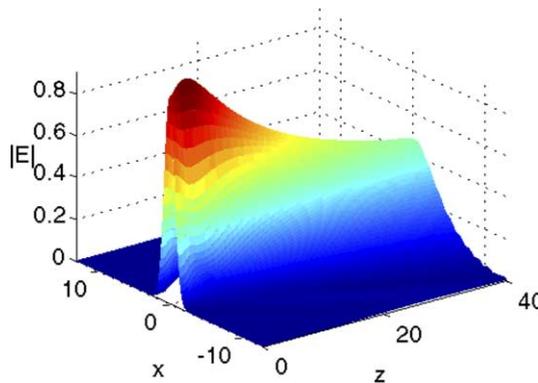


Fig. 7. Solution of the critical NLH subject to the Sommerfeld transverse boundary conditions for $x_{\max} = 13$.

sequence of grids ranging from twice as coarse, $(h_z, h_x) = (\lambda_0/10, \lambda_0/4)$, to twice as fine, $(h_z, h_x) = (\lambda_0/40, \lambda_0/16)$ compared with the original grid. Between each two consecutive grids in the sequence, the resolution changes by a factor of 2. The results are summarized in Table 1, where we present the maximum norms of the difference between consecutive solutions. Note that the improvement by reducing the grid sizes by a factor of two is $0.125/0.0102 \approx 12.25$, which is close to the theoretical prediction $2^4 = 16$.

Table 1
Grid convergence study for the Sommerfeld boundary conditions

Grid no.	$i = 1$	$i = 2$	$i = 3$
(h_z, h_x)	$(\frac{z_0}{10}, \frac{z_0}{4})$	$(\frac{z_0}{20}, \frac{z_0}{8})$	$(\frac{z_0}{40}, \frac{z_0}{16})$
$\ E^{(\text{grid} > i)} - E^{(\text{grid} > i-1)}\ _\infty$		0.125	0.0102

To further demonstrate the superiority of the new approach, let us now reduce the transverse size of the computational domain by approximately a factor of 2, to $x_{\max} = 7$. As we have already mentioned, for sizes smaller than 13 the iterations with the Dirichlet boundary conditions diverge. In contrast, iterations with the Sommerfeld boundary conditions still converge for $x_{\max} = 7$ at the same rate as for $x_{\max} = 13$, see Fig. 8. The quality of the Sommerfeld solution on the 40×7 domain, see Figs. 9 and 10 is only marginally worse than that for the 40×13 domain, and is roughly comparable with the quality of the 40×13 Dirichlet solution, see Figs. 3 and 4. In fact, the Sommerfeld boundary conditions allow us to reduce the transverse size of the domain even further and keep the same convergence rate up to at least $x_{\max} = 3$. However, for such a small domain the solution itself bears little resemblance to the solution which we are trying to approximate (which corresponds to the unbounded transverse domain), and so we do not present it here.

Going in the “opposite direction”, we started with the Sommerfeld solution for $x_{\max} = 13$ and tried to obtain a solution with comparable quality with the Dirichlet boundary conditions. In that case we have found out that the transverse size of the domain should roughly be twice as large, see Figs. 11–13 which show the results for the Dirichlet case with $x_{\max} = 25$.

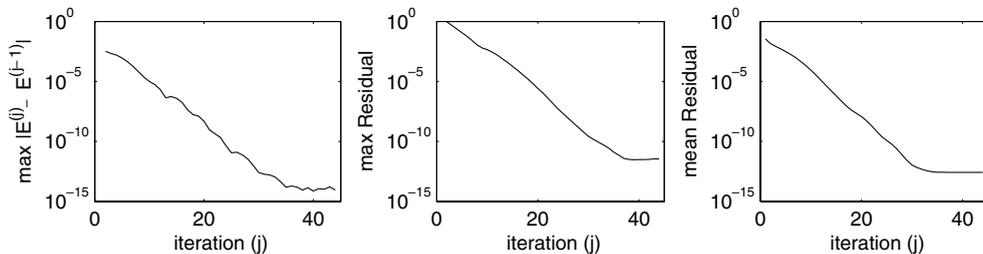


Fig. 8. Convergence history plots for the critical NLH subject to the Sommerfeld transverse boundary conditions; $x_{\max} = 7$.

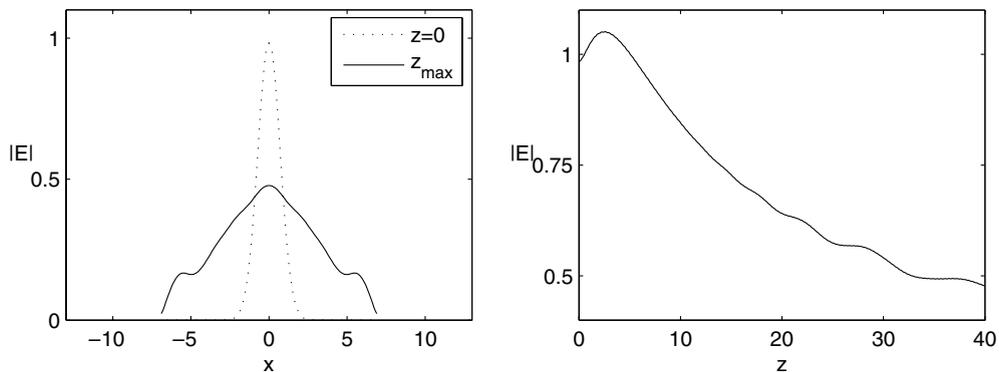


Fig. 9. Solution of the critical NLH subject to the Sommerfeld transverse boundary conditions $x_{\max} = 7$: Incoming and outgoing cross-sections (left) and on-axis amplitude $|E(0,z)|$ (right).

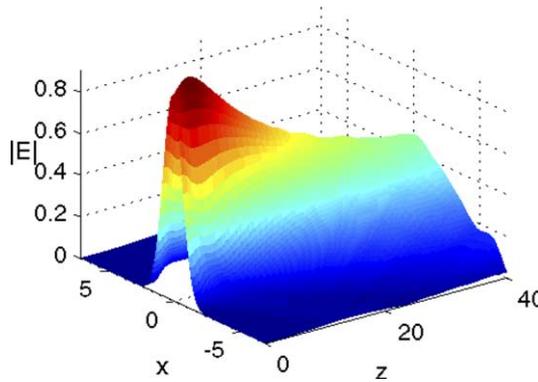


Fig. 10. Solution of the critical NLH subject to the Sommerfeld transverse boundary conditions for $x_{\max} = 7$.

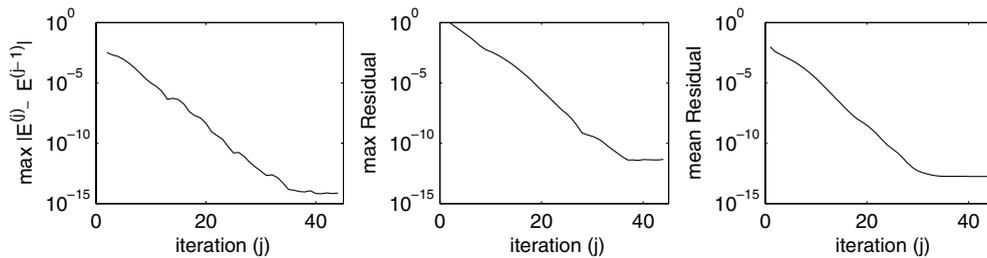


Fig. 11. Convergence history plots for the critical NLH subject to the Dirichlet transverse boundary conditions; $x_{\max} = 25$.

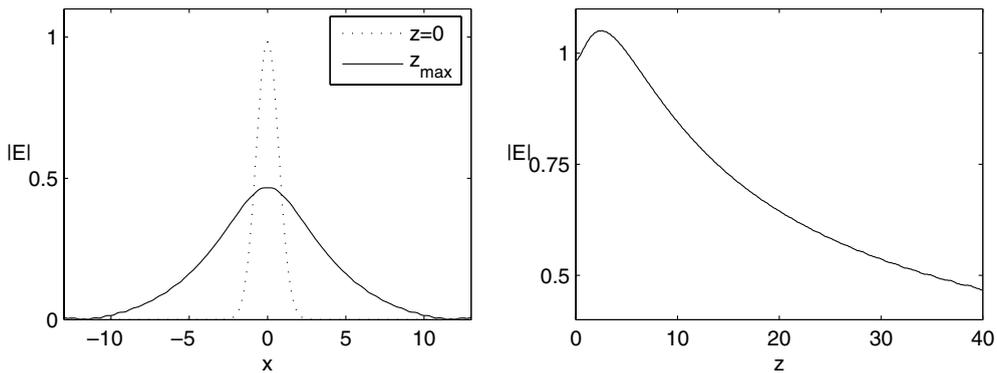


Fig. 12. Solution of the critical NLH subject to the Dirichlet transverse boundary conditions for $x_{\max} = 25$: Incoming and outgoing cross-sections (left) and on-axis amplitude $|E(0,z)|$ (right).

Finally, in Fig. 14 we show the results that correspond to higher input powers and as such, allow one to observe stronger self-focusing. These results are obtained with the Sommerfeld transverse boundary conditions: For $k_0 = 8$ on the domain $z_{\max} = 60$, $x_{\max} = 20$, $h_z = \lambda_0/20$, and $h_x = \lambda_0/8$ (Fig. 14(left)), and for $k_0 = 30$ on the domain $z_{\max} = 60$, $x_{\max} = 15$, $h_z = \lambda_0/5$, and $h_x = \lambda_0/4$ (Fig. 14(right)). In the computation with $p = 99.6\%$ the plot of the solution looks more stretched in the longitudinal direction, because we used

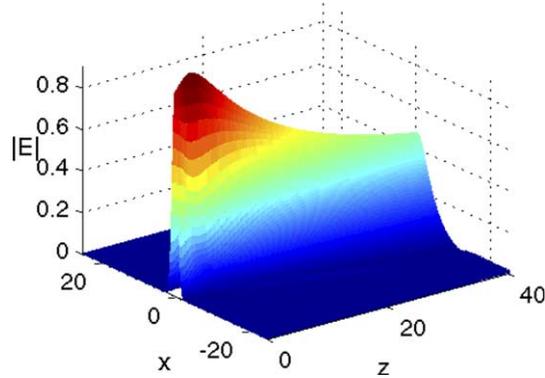


Fig. 13. Solution of the critical NLH subject to the Dirichlet transverse boundary conditions for $x_{\max} = 25$.

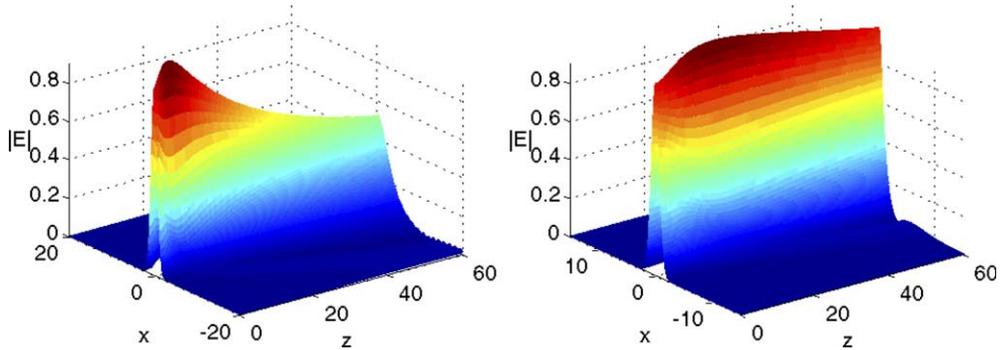


Fig. 14. Solution of the critical NLH subject to the Sommerfeld transverse boundary conditions for $k_0 = 8$ and $p = 99\%$ (left), and $k_0 = 30$ and $p = 99.6\%$ (right).

$k_0 = 30$, which implies a larger diffraction length $L_{\text{DF}} = k_0 r_0^2$. Also notice that solutions with such high input powers are very difficult and very expensive to obtain with the Dirichlet transverse boundary conditions.

To summarize, the Sommerfeld boundary conditions allow us to reduce the transverse size of the computational domain by roughly a factor of two without any degradation in either the solution quality or the convergence rate. This reduction also translates into a reduction of the computational costs. The translation, however, is not apparent and requires an additional comment. In the case of the Dirichlet boundary conditions, the diagonalization of system (54) that yields system (55) could be achieved by Fourier transform implemented via FFT, with a slightly super-linear asymptotic cost of $O(M \ln M)$ for each $n = 0, \dots, N$. With the Sommerfeld boundary conditions, however, the cost of diagonalization by means of the matrix Ψ of (52) is quadratic, $O(M^2)$ for each n . This may first seem a disadvantage of the new procedure. We need to realize however, that the physical problems that we consider, namely, propagation of beams essentially in the z direction, require neither large transverse domain sizes x_{\max} nor large transverse grid dimensions M , as long as one can guarantee (reasonably) good non-reflecting properties of the lateral boundaries $x = \pm x_{\max}$. For the moderate values of M that we have actually used, the overall computational cost is dominated by other parts of the algorithm, primarily, by the sparse linear solver applied to each system (56), rather than by the separation of variables. In Table 2, we present typical data of the relative cost

Table 2
Relative cost of the separation of variables

	Dirichlet BCs		Sommerfeld BCs	
	Domain size $z_{\max} \times x_{\max}$	40×13	40×25	40×6.5
Grid dimension $N \times M$	1019×132	1019×254	1019×66	1019×132
Percentage of overall runtime	6.45%	9.6%	7.5%	11%

of the separation of variable as percentage of the total execution time; the data are obtained by profiling the Matlab code.

Even though, as expected, the relative cost of the separation of variables is higher for the new algorithm, see Table 2, its contribution to the overall cost is insignificant in any case. If, on the other hand, there is ever need in considering larger transverse dimensions, a faster multiresolution-based procedure [23] can be employed that combines both the separation of variables and linear solver “under one roof.”

4.2. Subcritical case

We recall that the subcritical NLS (5) for $\sigma = 1$ admits a well-known soliton solution

$$A = \frac{e^{iz/2}}{\cosh(x/\sqrt{2})}.$$

To see how this soliton would propagate in the framework of the nonparaxial Helmholtz equation, we solve the two-dimensional NLH (6) with $\sigma = 1$ and with the incoming beam specified as

$$E^{(\text{inc, left})}(x) = \frac{(\epsilon k_0^2)^{-1/2}}{\cosh(x/\sqrt{2})}.$$

To maximize the effect of nonparaxiality and backscattering, we set $k_0 = 5$. As we can say that $r_0 \approx \sqrt{2}$, the diffraction length is $L_{\text{DF}} = k_0 r_0^2 \approx 10$. We also set $z_{\max} = 70$, which corresponds to propagation over 7 diffraction lengths.

The convergence history plots for the case $x_{\max} = 10$ can be seen in Fig. 15. For the Dirichlet transverse boundary conditions, iterations (23) quickly diverged. In contrast, the Sommerfeld boundary conditions have provided for a robust convergence at a linear rate; with the residuals dropping by about 9 orders of magnitude after 200 nonlinear cycles.

In order to better understand the aforementioned quick divergence of the iterations (23) for the Dirichlet boundary conditions, we plot in Fig. 16 the magnitude of the transformed solution $\hat{\mathcal{E}}_n = \Psi^{-1} \mathcal{E}_n = [\hat{E}_{0,n}, \dots, \hat{E}_{l,n}, \dots, \hat{E}_{2M,n}]$ for $n = 1$ as a function of $\text{Re } \alpha_z^2(l)$, see (63). In the Dirichlet case, Ψ^{-1} is the inverse discrete Fourier transform, $\hat{E}_{l,n}$ are Fourier coefficients, and α_z are real, hence $\text{Re } \alpha_z^2 = \alpha_z^2$. In the Sommerfeld case, $\hat{E}_{l,n}$ can no longer be referred to as genuine Fourier coefficients, and the quantities α_z are always separated from zero because $\text{Im } \lambda^{(l)} \neq 0$. Since $\text{Re } \alpha_z^2 < 0$ corresponds to the evanescent modes for both cases, we can expect most of the energy to remain concentrated in the region $\text{Re } \alpha_z^2 > 0$. This is, indeed, the case when we use the Sommerfeld transverse boundary conditions, see Fig. 16. In the case of the Dirichlet boundary conditions, however, we observe a rapid amplification of the near-zero modes ($|\alpha_z| \approx 0$). This numerical resonance is a manifestation of the computational instability due to degeneration of the one-way discrete Helmholtz equations mentioned in Section 3.8 – relations (64) and (65) become truly linearly dependent or “almost” linearly dependent for $\alpha_z = 0$ or $|\alpha_z| \ll 1$, respectively. In other words, the near-zero modes become poorly controlled by the

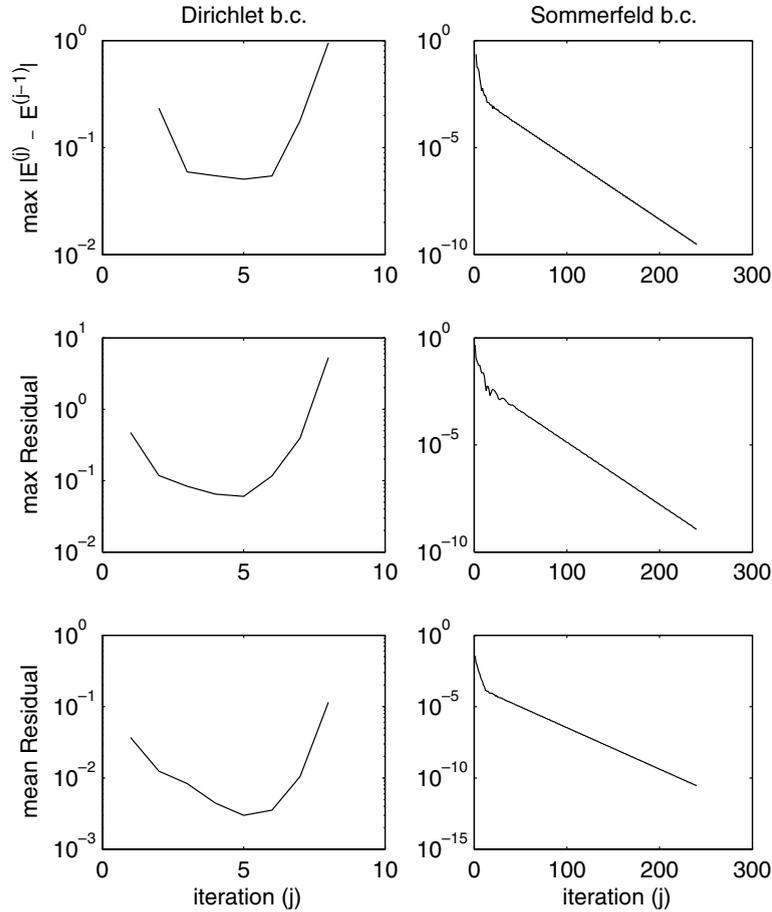


Fig. 15. Convergence histories for the case of a single soliton.

boundary conditions – in the ultimate case of the exact equality $\alpha_z = 0$ there is no control at all, and the homogeneous system admits a non-trivial solution.⁸ In the case of the Sommerfeld transverse boundary conditions, the modes for which $\text{Re } \alpha_z^2 \approx 0$ are the ones for which $\text{Re } \lambda^{(l)} \approx -k_0^2$, i.e., they correspond to the waves that basically propagate in the $\pm x$ directions. As predicted by the intuitive argument following Eq. (17), these waves would experience the largest amount of damping (see Fig. 17), which prohibits the growth of these modes. We thus see that the use of the Sommerfeld transverse boundary conditions removes the numerical resonance problem that may “haunt” the algorithm with the Dirichlet transverse boundary conditions.

In Fig. 18, we plot the transverse profiles of the solution at $z = 0$ and at $z = z_{\max}$. The discrepancy between the two is minute, which indicates that the NLH solution basically propagates as a soliton. The three-dimensional mesh plot of the solution in Fig. 19 corroborates the same observation. To

⁸ Whether the instability would manifest itself in each particular computation apparently depends on whether the nonlinear iterations (23) can pump more and more energy into the unstable modes.

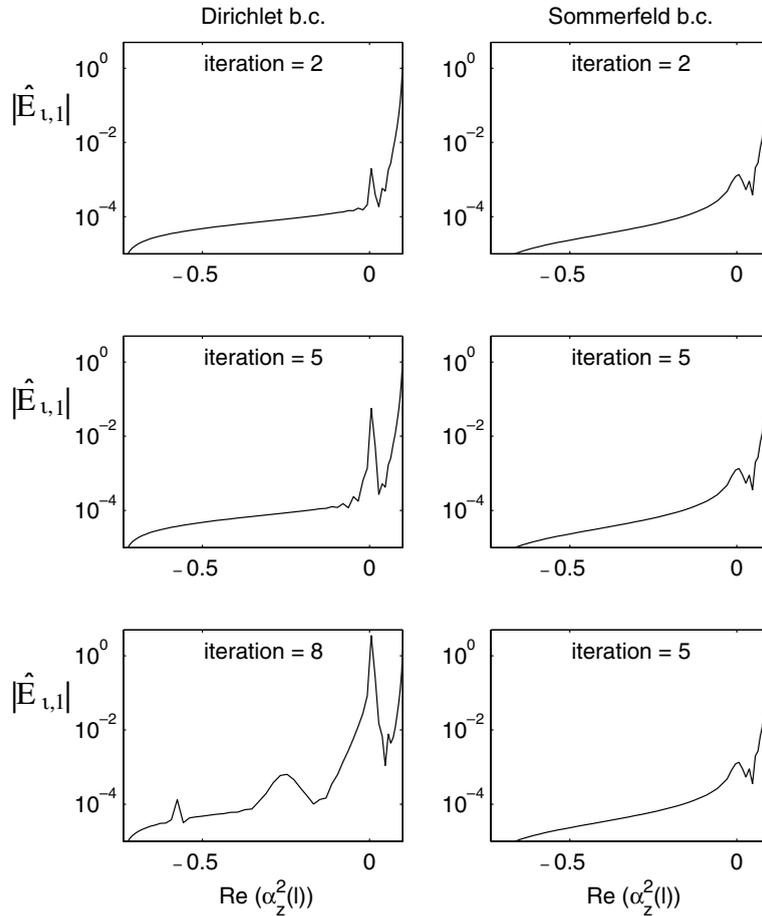


Fig. 16. Resonance monitoring for the case of a single soliton.

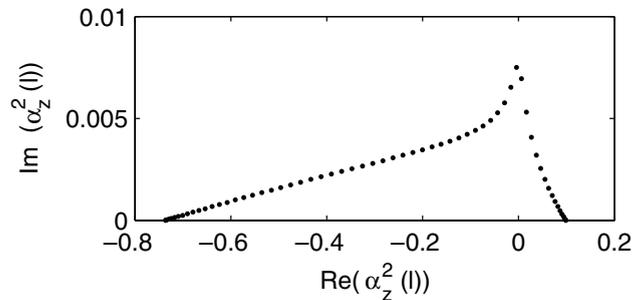


Fig. 17. Real and imaginary parts of $\alpha_z^2(l)$ for the simulation of a single soliton with the Sommerfeld boundary conditions.

the best of our knowledge, this is the first numerical evidence that *nonparaxiality and backscattering have a negligible effect on the propagation of solitons*, even when the soliton width is of the order of a single wavelength.

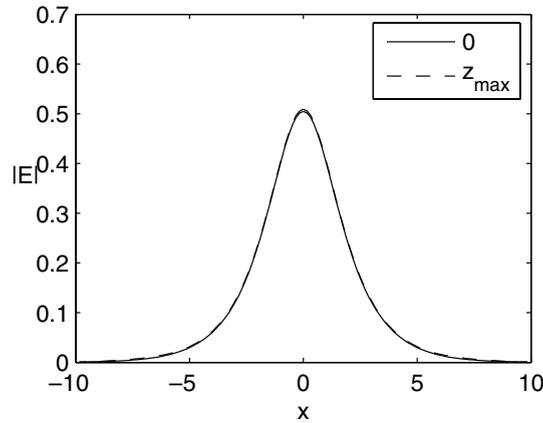


Fig. 18. Incoming and outgoing cross-sections of the single soliton solution to the subcritical NLH.

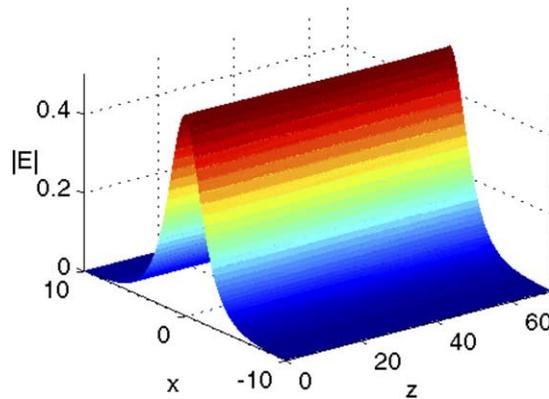


Fig. 19. Single soliton solution to the subcritical NLH.

In order to see how well our boundary conditions can treat the incoming waves that are not parallel to the z axis, we solve the subcritical NLH with the incoming beam

$$E^{(\text{inc, left})}(x, 0) = \frac{\exp(-ik_0 \sin \theta(x-5))(\epsilon k_0^2)^{-1/2}}{\cosh((x-5) \cos \theta / \sqrt{2})}, \quad \theta = \arctan(1/4),$$

which corresponds to a solitonic initial condition impinging on the $z = 0$ interface at an angle $\theta \neq 0^\circ$. The numerical parameters are as before, except that we set $x_{\text{max}} = 15$. As can be seen in Figs. 20–22, the solitonic propagation is unaffected by the non-normal incidence angle. In particular, there is no indication of spurious numerical reflections from the artificial boundaries, showing that our algorithm is capable of handling well the waves not aligned with the grid/boundaries.

Another significant advantage of our NLH solver is *its natural ability to solve*, apparently for the first time ever, *for counter-propagating beams without making any additional assumptions*. This is in contrast with the NLS model that admits propagation in one direction only and as such would require

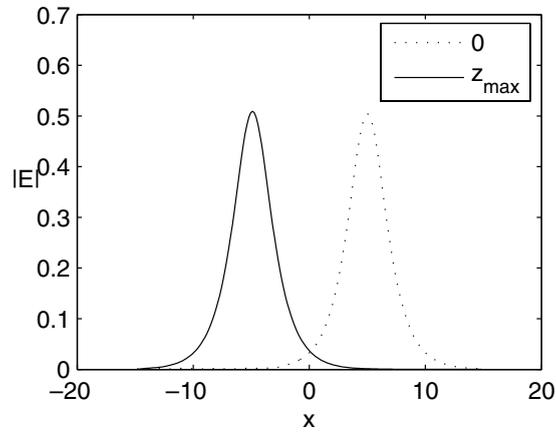


Fig. 20. Soliton at an angle: Cross-section profiles at $z = 0$ and $z = z_{\max}$.

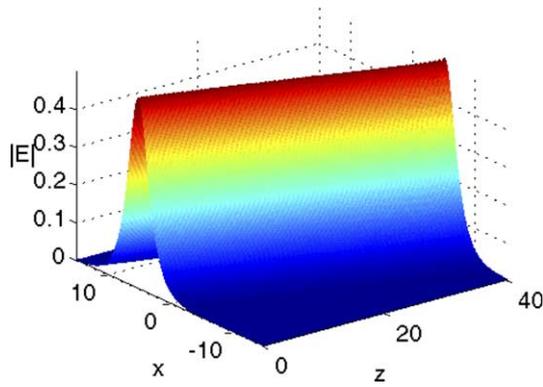


Fig. 21. Soliton at an angle: 3D mesh plot.

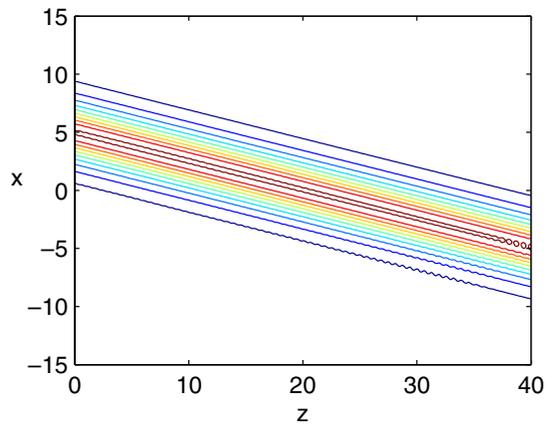


Fig. 22. Soliton at an angle: Level curves.

numerical marching in two opposite directions when analyzing the counter-propagation case. Note that although coupled NLS models have been used for counter-propagation, see [17], this approach involves some approximation of the Kerr nonlinearity induced by two coherent beams. In Figs. 23 and 24, we show our numerical solution of the subcritical NLH for the case of two counter-propagating incoming beams:

$$E^{(\text{inc,left})}(x, 0) = E^{(\text{inc,right})}(x, z_{\text{max}}) = \frac{\exp(-ik_0 \sin \theta(x - 10))(\epsilon k_0^2)^{-1/2}}{\cosh(\cos \theta(x - 10)/\sqrt{2})},$$

where $\theta = \pi/4$. The computational domain is $x_{\text{max}} = 20$, $z_{\text{max}} = 20$, and the grid sizes are $h_z = \lambda_0/60$ and $h_x = \lambda_0/16$. The centers of the two input beams are located at $(z = 0, x = 10)$ and $(z = 20, x = 10)$, the beams propagate downward at an angle of 45° , interact nonlinearly, and then continue to propagate as solitons. The periodic oscillations in the interaction region result from the phase differences between the two counter-propagating beams, they disappear as the beams move further away from each other.

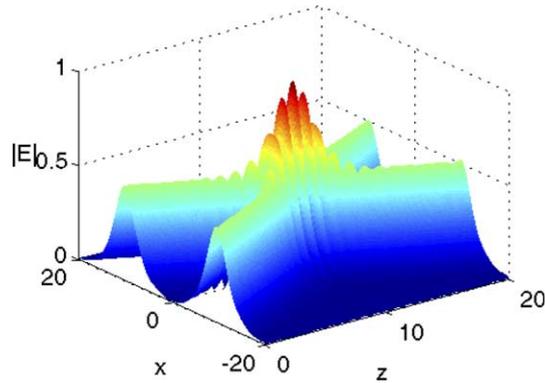


Fig. 23. Counter-propagation of two solitons at an angle: 3D mesh plot.

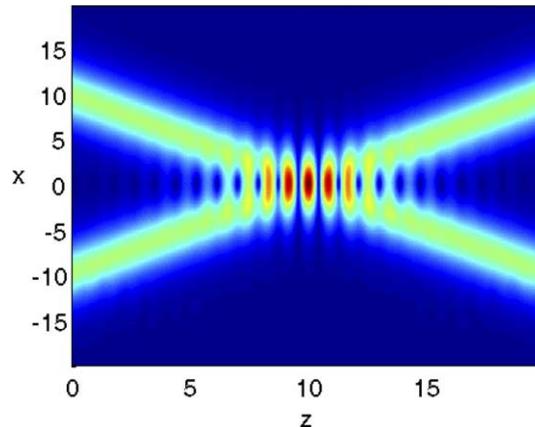


Fig. 24. Counter-propagation of two solitons at an angle.

5. Discussion

In this section, we outline our general motivation behind making a number of key choices in the design of the foregoing algorithm, present some elements of comparison to other available techniques, and finally draw our conclusions.

5.1. Sommerfeld boundary conditions

Boundary conditions (11) specify the desired outward direction of the radiation of waves at the lateral artificial boundaries. As such, they can naturally be called Sommerfeld radiation conditions, even though the original Sommerfeld conditions are typically set at infinity rather than at a finite boundary [24]. It is well known [5] that the fundamental limitation of boundary conditions (11) is that they are basically one-dimensional. In other words, any wave traveling toward either of the boundaries $x = x_{\max}$ or $x = -x_{\max}$ with non-normal incidence will only partially get through while the other part will get reflected back into the domain. In general, such reflections should be minimized, and ideally, we would, of course, like to have a boundary that ensures reflectionless propagation for all the outgoing waves. This, however, is only possible if the corresponding ABC is nonlocal, see [5]; and in the current paper (see also [6–8]) the nonlocal treatment is reserved for the artificial boundaries $z = 0$ and $z = z_{\max}$. As for the lateral boundaries $x = x_{\max}$ and $x = -x_{\max}$, the next best choice after the previously used Dirichlet conditions [6–8], is obviously the local radiation conditions (11).

Note that a number of different methods have been proposed in the literature, see the review [5], aimed at better absorbing the outgoing radiation with non-normal incidence. They can be said to occupy an intermediate position between the fully nonlocal ABCs and the local Sommerfeld conditions of type (11). In this category, one should first mention higher order local conditions by Bayliss and Turkel [25–27], as well as those by Higdon [28–33]. The first-order Sommerfeld conditions (11) can then be interpreted as the first member of the Bayliss–Turkel hierarchical sequence of boundary conditions. Other members of this sequence have been shown to make the artificial boundary more transparent for the outgoing waves with non-normal incidence. However, they also involve differentiation of orders higher than one, cf. formulae (11), which clearly brings along the issue of accurate and stable discrete approximation. Therefore, in the current paper we have only employed the first-order local radiation boundary conditions in the transverse direction. Instead of pursuing continuous higher orders, say, in the sense of [25–27], we rather focus on a very thorough implementation of boundary conditions (11) in the context of a fourth-order finite-difference approximation, see Section 3. Compared to the previously used Dirichlet boundary conditions [6], this modification alone already provides for very substantial computational benefits, as the computations of Section 4 show. At the same time, we admit that even better results could have possibly been obtained if a continuous higher-order boundary condition were to be implemented as carefully in the discrete context as we have implemented conditions (11). We plan to study this question in the future.

Another potential alternative to the transverse boundary conditions of type (11) is offered by perfectly matched layers (PML). A large number of such techniques have been proposed in the literature since the original work [34,35]; a particular example for time-harmonic waves can be found in [36]. It is generally acknowledged that PMLs have better absorption characteristics for the outgoing waves than local radiation boundary conditions do. In the context of the current methodology, however, while it is important to facilitate the reflectionless propagation of waves through the lateral boundaries $x = \pm x_{\max}$, it is even more important to be able to carry out the separation of variables described in Section 2.2, as otherwise the crucial nonlocal two-way ABCs at $z = 0$ and $z = z_{\max}$ cannot be set. The separation of variables is rendered by the expansion with respect to the transverse eigenfunctions. However, to the best of our knowledge, nobody has ever looked into the spectra and eigenfunctions of the Helmholtz operators on the domains that include

regions of PML (variable coefficients). The corresponding analysis may prove cumbersome, and is certainly beyond the scope of the current paper. We therefore, again, leave it for a future study.

Another important comment is in order. Boundary conditions (11) are derived in the framework of the linear equation (9). The reason is that the geometry of the problem (see Fig. 1) suggests that if the lateral artificial boundaries are located sufficiently far from the central part of the domain where the nonlinear phenomena take place, then the nonlinearity near $x = x_{\max}$ and $x = -x_{\max}$ can merely be disregarded. We realize, of course, that this argument is only qualitative and does not, for example, allow one to judge ahead of time how far away those lateral boundaries shall be placed. Still, numerical experiments of Section 4 do show that they can be located much closer for the Sommerfeld conditions than for the previously used Dirichlet conditions, and that overall the linear boundary conditions (11) (more precisely, discrete boundary conditions (39)) have performed exceptionally well, which provides a conclusive justification of their legitimacy.

In the Literature, one can find considerations similar to those that lead to the continuous homogeneous boundary conditions (11) and (18), as well as to the inhomogeneous two-way boundary conditions (20), see, e.g., [37]. However, the corresponding discrete boundary conditions (39), (64) and (65), as well as (71) and (72), are unparalleled, as they are obtained by the direct analysis of the waves propagating on the grid in the context of a fourth-order discretization, as opposed, e.g., to asymmetric one-sided differencing.

5.2. Local versus nonlocal boundary conditions

In spite of their apparent similarity, an important distinction is to be emphasized between boundary conditions (11) and boundary conditions (18). The Sommerfeld conditions (11) can only guarantee the reflectionless propagation for the waves that have normal incidence. As such, in the two-dimensional setting (x, z) these boundary conditions appear approximate. Conversely, as each equation (17) is genuinely one-dimensional, the corresponding boundary conditions (18) are exact. Indeed, in the one-dimensional case there may be no other angle of incidence for the outgoing waves except normal. Of course, this exactness is only enabled by the analysis conducted in the transformed space, i.e., after the second dimension x has been “removed” by the separation of variables and replaced by the index l . We note that for different l s the wavenumbers $\sqrt{k_0^2 + \lambda^{(l)}}$ of the waves propagating in the z direction are also different, as they depend on the eigenvalue $\lambda^{(l)}$. To maintain exactness of the ABCs in the original two-dimensional space, we need to transform back from the coefficients $\hat{E}_l(z)$ to the functions $E(x, z)$. This is a fairly standard approach to obtaining the ABCs, see the review [5]. In practically all similar cases reported in the literature, it yields nonlocal pseudodifferential operators instead of the symbols $\pm i\sqrt{k_0^2 + \lambda^{(l)}}$ in formulae (18). In general, the nonlocality of the ABCs is “the price to pay” for their exactness, i.e., for the capability to accurately handle the entire range of the angles of incidence in the (x, z) setting, which, as one can easily see, is equivalent to accommodating different wavenumbers for different l s in the transformed space. In the literature, the nonlocality of the ABCs has often been considered a disadvantage and a potential source of difficulties at the implementation stage. Contrary to this “conventional wisdom,” it turns out that in many cases nonlocal ABCs can be efficiently implemented in the discrete framework [5], which typically results in algorithms with superb performance. In particular, in the current paper we have seen that there was no need to explicitly transform the ABCs back into the physical space, because they are naturally implemented in the context of a solver based on the separation of variables, see Section 3.7.

6. Conclusions

Altogether, the Sommerfeld transverse boundary conditions render a considerably more robust computational procedure, compared to the previously used Dirichlet boundary conditions. On the computational

domain of the same size, the quality of the solution obtained with the Sommerfeld boundary conditions is noticeably higher than that for the Dirichlet boundary conditions. Besides, due to the reduced reflections from the lateral boundaries, the Sommerfeld boundary conditions facilitate convergence of the iteration scheme (23) on much smaller domains than the Dirichlet boundary conditions can handle.

Moreover, the new algorithm brings along new computational capabilities that could not be attained previously. In particular, it offers a natural way of solving for counter-propagating solitons (see Section 4.2) in the framework of a single boundary value problem for the NLH with no approximations needed. For the same setup, the NLS model would require marching two Cauchy problems in the opposite directions with only approximate treatment of the nonlinear interaction between the two beams.

Recently, it has also been brought to our attention that the properties of the Sommerfeld eigenvectors established in Section 3, in particular, their linear independence (Proposition 7), as well as the easy way to invert the transformation Ψ according to formula (53), have been observed experimentally in the context of building the direct solve preconditioners for large-scale scattering problems, see, e.g., [38,39]. However, to the best of our knowledge no justification of these properties similar to the one given in Section 3 has ever been provided, and we therefore expect that our results may appear useful not only for solving the NLH, but for other applications as well.

Finally, solving the critical NLH for input powers above critical still remains an outstanding difficult issue. Its analysis may require a major new insight, and will be a subject of our future work.

Acknowledgements

We thank Michael Weinstein for suggesting the use of non-self-adjoint transverse boundary conditions, and for bringing Ref. [19] to our attention. The Research of G. Fibich was partially supported by Grant No. 2000311 from the United States–Israel Binational Science Foundation (BSF), Jerusalem, Israel.

References

- [1] R. Boyd, *Nonlinear Optics*, 2nd edition., Academic Press, Boston, 2003.
- [2] C. Sulem, P.L. Sulem, *The Nonlinear Schrödinger Equation*, Springer, New York, 1999.
- [3] G. Fibich, G.C. Papanicolaou, Self-focusing in the perturbed and unperturbed nonlinear Schrödinger equation in critical dimension, *SIAM J. Appl. Math.* 60 (1999) 183–240.
- [4] L. Wöste, C. Wedekind, H. Wille, P. Rairoux, B. Stein, S. Nikolov, C. Werner, S. Niedermeier, F. Ronnenberger, H. Schillinger, R. Sauerbrey, Femtosecond atmospheric lamp, *Laser und Optoelektronik* 29 (1997) 51–53.
- [5] S.V. Tsynkov, Numerical solution of problems on unbounded domains. A review, *Appl. Numer. Math.* 27 (1998) 465–532.
- [6] G. Fibich, S.V. Tsynkov, High-order two-way artificial boundary conditions for nonlinear wave propagation with backscattering, *J. Comput. Phys.* 171 (2001) 632–677.
- [7] G. Fibich, B. Ilan, S.V. Tsynkov, Computation of nonlinear backscattering using a high-order numerical method, *J. Sci. Comput.* 17 (1–4) (2002) 351–364.
- [8] G. Fibich, B. Ilan, S.V. Tsynkov, Backscattering and nonparaxiality arrest collapse of damped linear waves, *SIAM J. Appl. Math.* 63 (5) (2003) 1718–1736.
- [9] S.V. Tsynkov, External boundary conditions for three-dimensional problems of computational aerodynamics, *SIAM J. Sci. Comp.* 21 (1999) 166–206.
- [10] G. Fibich, B. Ilan, Vectorial and random effects in self-focusing and in multiple filamentation, *Physica D* 157 (2001) 112–146.
- [11] M.I. Weinstein, Nonlinear Schrödinger equations and sharp interpolation estimates, *Comm. Math. Phys.* 87 (1983) 567–576.
- [12] N. Akhmediev, J.M. Soto-Crespo, Generation of a train of three-dimensional optical solitons in a self-focusing medium, *Phys. Rev. A* 47 (1993) 1358–1364.
- [13] N. Akhmediev, A. Ankiewicz, J.M. Soto-Crespo, Does the nonlinear Schrödinger equation correctly describe beam propagation? *Opt. Lett.* 18 (1993) 411–413.
- [14] M.D. Feit, J.A. Fleck, Beam nonparaxiality, filament formation, and beam breakup in the self-focusing of optical beams, *J. Opt. Soc. Am. B* 5 (1988) 633–640.

- [15] G. Fibich, Small beam nonparaxiality arrests self-focusing of optical beams, *Phys. Rev. Lett.* 76 (1996) 4356–4359.
- [16] M. Sever, An existence theorem for some semilinear elliptic systems, *J. Differential Eqs.*, submitted for publication.
- [17] O. Cohen, R. Uzdin, T. Carmon, J. Fleischer, M. Segev, S. Odoulov, Collisions between optical spatial solitons propagating in opposite directions, *Phys. Rev. Lett.* 89 (2002) 133901.
- [18] M.S. Agranovich, B.Z. Katsenelenbaum, A.N. Sivov, N.N. Voitovich, *Generalized Method of Eigenoscillations in Diffraction Theory*, WILEY-VCH Verlag Berlin GmbH, Berlin, 1999, translated from the Russian manuscript by Vladimir Nazaikinskii.
- [19] E.S.C. Ching, P.T. Leung, A.M. van der Brink, W.M. Suen, S.S. Tong, K. Young, Quasi-normal mode expansion for waves in open systems, *Rev. Modern Phys.* 70 (4) (1998) 1545–1554.
- [20] M.S. Agranovich, Elliptic boundary problems, in: *Partial Differential Equations, IX Encyclopaedia of Mathematical Science*, vol. 79, Springer, Berlin, 1997, pp. 1–144, 275–281, translated from the Russian by the author.
- [21] V.S. Ryaben’kii, Necessary and sufficient conditions for good definition of boundary value problems for systems of ordinary difference equations, *U.S.S.R. Comput. Math. Math. Phys.* 4 (1964) 43–61.
- [22] S.K. Godunov, V.S. Ryaben’kii, *Difference Schemes: an Introduction to Underlying Theory*, North-Holland, New York, Amsterdam, 1987.
- [23] T. Rossi, J. Toivanen, A nonstandard cyclic reduction method, its variants and stability, *SIAM J. Matrix Anal. Appl.* 20 (3) (1999) 628–645 (electronic).
- [24] R. Courant, D. Hilbert *Methods of Mathematical Physics*, vol. II, Wiley, New York, 1962.
- [25] A. Bayliss, E. Turkel, Radiation boundary conditions for wave-like equations, *Comm. Pure Appl. Math.* 33 (6) (1980) 707–725.
- [26] A. Bayliss, E. Turkel, Outflow boundary conditions for fluid dynamics, *SIAM J. Sci. Statist. Comput.* 3 (2) (1982) 250–259.
- [27] A. Bayliss, E. Turkel, Far field boundary conditions for compressible flows, *J. Comput. Phys.* 48 (2) (1982) 182–199.
- [28] R.L. Higdon, Radiation boundary conditions for dispersive waves, *SIAM J. Numer. Anal.* 31 (1) (1994) 64–100.
- [29] R.L. Higdon, Absorbing boundary conditions for difference approximations to the multidimensional wave equation, *Math. Comp.* 47 (176) (1986) 437–459.
- [30] R.L. Higdon, Absorbing boundary conditions for acoustic and elastic waves in stratified media, *J. Comput. Phys.* 101 (2) (1992) 386–418.
- [31] R.L. Higdon, Absorbing boundary conditions for dispersive waves, in: T.L. Geers (Ed.), *IUTAM Symposium on Computational Methods for Unbounded Domains* (Boulder, CO, 1997), *Fluid Mechanics and Applications*, vol. 49, Kluwer Acad. Publ., Dordrecht, 1998, pp. 217–220.
- [32] R.L. Higdon, Numerical absorbing boundary conditions for the wave equation, *Math. Comp.* 49 (179) (1987) 65–90.
- [33] R.L. Higdon, Radiation boundary conditions for elastic wave propagation, *SIAM J. Numer. Anal.* 27 (4) (1990) 831–869.
- [34] J.-P. Berenger, A perfectly matched layer for the absorption of electromagnetic waves, *J. Comput. Phys.* 114 (2) (1994) 185–200.
- [35] J.-P. Berenger, Three-dimensional perfectly matched layer for the absorption of electromagnetic waves, *J. Comput. Phys.* 127 (2) (1996) 363–379.
- [36] S.V. Tsynkov, E. Turkel, A cartesian perfectly matched layer for the Helmholtz equation, in: L. Tournette, L. Halpern (Eds.), *Absorbing Boundaries and Layers, Domain Decomposition Methods. Applications to Large Scale Computations*, Nova Science Publishers, Inc., New York, 2001, pp. 279–309.
- [37] F.B. Jensen, W.A. Kuperman, M.B. Porter, H. Schmidt, *Computational ocean acoustics* AIP Series in Modern Acoustics and Signal Processing, American Institute of Physics, New York, 1994.
- [38] T. Rossi, J. Toivanen, A parallel fast direct solver for block tridiagonal systems with separable matrices of arbitrary dimension, *SIAM J. Sci. Comput.* 20 (5) (1999) 1778–1796 (electronic).
- [39] E. Heikkola, T. Rossi, J. Toivanen, Fast direct solution of the Helmholtz equation with a perfectly matched layer or an absorbing boundary condition, *Int. J. Numer. Methods Engrg.* 57 (14) (2003) 2007–2025.