BIT manuscript No.
(will be inserted by the editor)

Analysis of a Fast Method for Solving the High Frequency Helmholtz Equation in One Dimension

Jelena Popovic · **Olof Runborg**

Received: date / Accepted: date

Abstract We propose and analyze a fast method for computing the solution of the high frequency Helmholtz equation in a bounded one-dimensional domain with a variable wave speed function. The method is based on wave splitting. The Helmholtz equation is split into one–way wave equations with source functions which are solved iteratively for a given tolerance. The source functions depend on the wave speed function and on the solutions of the one-way wave equations from the previous iteration. The solution of the Helmholtz equation is then approximated by the sum of the one-way solutions at every iteration. To improve the computational cost, the source functions are thresholded and in the domain where they are equal to zero, the one-way wave equations are solved with geometrical optics with a computational cost independent of the frequency. Elsewhere, the equations are fully resolved with a Runge–Kutta method. We have been able to show rigorously in one dimension that the algorithm is convergent and that for fixed accuracy, the computational cost is asymptotically just $\mathcal{O}(\omega^{1/p})$ for a p-th order Runge–Kutta method, where ω is the frequency. Numerical experiments indicate that the growth rate of the computational cost is much slower than a direct method and can be close to the asymptotic rate.

Keywords Helmholtz equation · High frequency · Wave splitting

Mathematics Subject Classification (2000) 65N06 · 35L05

1 Introduction

Simulation of high frequency wave propagation is important in many engineering and scientific disciplines. Currently the interest is driven by new applications in wireless communication (cell phones, bluetooth) and photonics (optical fibers, filters, switches). Simulation is also used in more classical applications. Some examples in electromagnetics are antenna

J. Popovic
Department of Numerical Analysis, CSC, KTH, 100 44 Stockholm
E-mail: jelenap@csc.kth.se

O. Runborg
Department of Numerical Analysis, CSC, KTH and Swedish e-Science Research Center (SeRC), 100 44 Stockholm
E-mail: olofr@csc.kth.se

design and radar signature computation. In acoustics simulation is used for noise prediction, underwater communication and medical ultrasonography. In this paper we consider wave propagation problems that are governed by the 1D Helmholtz equation

$$u_{xx} + \frac{\omega^2}{c(x)^2} u = 0, \qquad \omega \gg 1, \qquad x \in (-L, L) \tag{1}$$

where $c(x)$ is the speed of propagation, and ω is the frequency. The equation is augmented by suitable boundary conditions at $x = \pm L$.

One significant difficulty with direct numerical simulation of high frequency wave problems is that the wavelength is very short compared to the computational domain and thus many discretization points are needed to resolve the solution. For the Helmholtz equation standard discretizations by finite difference or finite element schemes leads to sparse systems of equations with a number of unknowns N that depends on the frequency, $N \sim \omega^d$, where d is the dimension of the problem. Those large systems are usually solved by iterative methods. However, since the system of equations is indefinite and ill-conditioned, the convergence rate can be very slow. Finding good preconditioners for Helmholtz is a major challenge, see e.g. [21,34,38,23,16] and more recently [22,20,19], but even with good precondioners the computational complexity is typically super linear in N. The cost therefore grows at least as ω^d and direct numerical methods cannot be used as $\omega \to \infty$. Hence, development of effective numerical methods for such problems is important.

An alternative approach is to use an asymptotic approximation such as geometrical optics (GO). See for example [17,46,44,6]. Instead of the oscillating wave field, the unknowns in GO are the phase and the amplitude which vary on a much coarser scale than the full solution. They are therefore easier to compute numerically, at a cost independent of the frequency. There are many numerical methods based on the GO approximation, see for example [13,18,24,32,43,45,47]. The main disadvantage of the GO approximation is that it is only accurate for large frequencies. It typically requires that variations in the speed of propagation $c(x)$ are on a scale much larger than the wave length. Another disadvantage of GO is that it cannot capture diffracted waves that are produced when the incident field hits edges, corners or vertexes of the obstacle or when the incident wave hits the tangent points of the smooth scatterer (creeping waves). Moreover, GO fails at caustics where waves focus. Techniques to overcome these problems include geometrical theory of diffraction [30] and Gaussian beams [40,42].

The situation described above can be summarized as follows: For direct methods the computational cost grows with the frequency for fixed accuracy, while for GO methods the accuracy grows with the frequency for fixed computational cost. Unfortunately, the frequency range and accuracy requirement of many realistic problems often fall in between what is tractable with either of these approaches. Recently a new class of algorithms has been proposed that combine the cost advantage of GO methods with the accuracy advantage of direct methods. They are thus characterized by a computational cost that grows slowly with the frequency, while at the same time being accurate also for moderately high frequencies.

The main interest of the new methods has been for scattering problems [8,31,14,1, 28,11]. Those methods are based on the integral formulation of the Helmholtz equation. The idea is to make an approximation space that incorporates the oscillatory behavior of the solution so that the new unknown function is less oscillatory than the original. Those methods require in principle only $\mathcal{O}(1)$ unknowns as $\omega \to \infty$ and they are also able to keep the computational cost at the same level using special numerical methods for oscillatory

integrals, see [31,8]. Most of this work has been done for convex scatterers but there are also extensions to non-convex scatterers [9,15]. Rigorous proofs of the low cost of these methods are difficult and require detailed results on the asymptotic behavior of the exact solution near shadow boundaries and corners. An example of a result, due to Dominguez, Graham and Smyshlyaev [14], concerns convex smooth scatterers in 2D. They show that the error is controlled if the number of unknowns N and computational cost are slightly larger than $k^{1/9}$. Another result is given by Chandler-Wilde and Langdon for convex polygons where, similarly, the cost only grows as $\log k$, see [12].

For full domain problems of the type (1) much less has been done. In some sense this is more difficult than the scattering problem in a homogeneous medium because the waves are reflected at all points where the wave speed changes, not only at the surface of a scatterer. One attempt at lowering the computational cost along the lines above has been to use *plane wave basis functions* in finite element methods [39,37,4,10]. The method can be seen as a discontinuous Galerkin method with a particular choice of basis functions and numerical flux [26,35]. However, except in simple cases these methods do not reduce the complexity more than by a constant factor. The method proposed by Giladi and Keller [25] is a hybrid numerical method for the Helmholtz equation in which the finite element method is combined with GO. The idea is to determine the phase factor which corresponds to the plane wave direction a priori by solving the eikonal equation for the phase using ray tracing and then to determine the amplitude by a finite element method choosing asymptotically derived basis functions which incorporate the phase factor. Han and Huang [27] recently proposed a *tailored finite-point method* for the 1D Helmholtz equation. They divide the domain into intervals, in which they approximate the wave speed function $c(x)$ by a piecewise constant function. On each interval they solve the equation exactly and couple the solutions via transmission and reflexion conditions. The speed function is assumed to be smooth and monotone in each interval. They analyze a case when the solution decays to zero with ω and show a frequency independent absolute (but not relative) error.

In this article we introduce a weakly frequency dependent method based on wave splitting for the 1D full domain problem (1). Although there are already some interesting applications in one dimension [33,3], most practical problems are of course in higher dimensions. However, our focus in this article is on the analysis and to show that, at least in one dimension, it is possible to construct a method for which one can show rigorous error and cost estimates that depend weakly on the frequency. The Helmholtz equation is split into one-way wave equations which are solved iteratively, sweeping back and forth over the domain. The full solution of (1) is approximated by the sum of those solutions. We show that this sum converges rapidly. The one-way equations have source functions that depend on the solutions of the one-way wave equations from the previous iteration. To improve the computational cost, the source functions are thresholded and in the domain where they are equal to zero, the one-way wave equations can be solved with geometrical optics with a computational cost independent of the frequency. Elsewhere, the equations are fully resolved. This is only necessary in a small domain of size $1/\omega$, however, and we are able to show that for fixed accuracy, the computational cost is asymptotically just $\mathcal{O}(\omega^{1/p})$ for a p-th order Runge–Kutta method. We also perform numerical experiments and the results indicate that in practice the cost can indeed be close to this.

The paper is organized as follows. In Section 2 we derive our method. In Section 3 we derive estimates of the solution to the one-way wave equation and show the convergence of the algorithm. In Section 4 we do the error analysis of the numerical implementation and show that the computational cost depends only weakly on the frequency. Numerical examples are given in Section 5.

2 A Fast Method for Helmholtz Equation

In this section we derive the fast method. We consider the 1D Helmholtz equation

$$u_{xx} + \frac{\omega^2}{c(x)^2} u = \omega f, \quad x \in (-L, L), \tag{2}$$

where ω is the frequency and $c(x)$ is the wave–speed function such that $\text{supp}(c_x) \subset (-L, L)$. It is augmented with the non-reflecting boundary conditions

$$u_x(-L) - i\omega u(-L) = -2i\omega A, \tag{3}$$

$$u_x(L) + i\omega u(L) = 0, \tag{4}$$

which also incorporate an incoming wave with amplitude A. We are mainly concerned with the homogeneous case when $f = 0$, but keep the more general form of the equation as it is needed in the subsequent analysis also for the $f = 0$ case. At high frequencies GO is a good approximation of the solution. We want to find a way to correct for the errors it makes at lower frequencies. A natural idea would be to use the system of WKB equations

$$\begin{aligned} 2\nabla\phi \cdot \nabla A_0 + \Delta\phi A_0 &= 0 \\ &\vdots \qquad\qquad |\nabla\phi| = \frac{1}{c} \\ 2\nabla\phi \cdot \nabla A_{n+1} + \Delta\phi A_{n+1} &= \Delta A_n, \end{aligned} \tag{5}$$

for the amplitude and phase, that is obtained in GO when the solution is approximated by

$$u(\mathbf{x}) = e^{i\omega\phi(\mathbf{x})} \sum_{k=0}^{\infty} A_k(\mathbf{x})(i\omega)^{-k}. \tag{6}$$

However, the series in (6) does not converge, even in simple settings. It is only an asymptotic series. The main problem is that (5) only describes waves traveling in one direction. In reality, waves are reflected whenever $c_x \neq 0$. We therefore introduce the functions v and w to describe waves propagating in the right-going and the left-going directions, respectively. The full solution is the sum of these functions, $u = v + w$. We make the ansatz that v and w satisfy the following one-way wave equations,

$$i\omega v + c(x)v_x - \frac{1}{2}c_x(x)v = F, \tag{7}$$

$$i\omega w - c(x)w_x + \frac{1}{2}c_x(x)w = F, \tag{8}$$

with F to be determined. If $z = v + w$, then from a simple manipulation of (7) and (8) it follows that z satisfies the following equation

$$c^2 z_{xx} + \omega^2 z = -2i\omega F + \alpha(x)z, \qquad \alpha(x) = \frac{1}{2}cc_{xx} - \frac{1}{4}c_x^2. \tag{9}$$

Thus, if $F = \alpha(x)z/2i\omega$, then $z = u$, the solution of the Helmholtz equation (2) with $f = 0$. We now make an expansion of v and w in powers of ω,

$$v = \sum_{n=0}^{\infty} r_n \omega^{-n}, \qquad w = \sum_{n=0}^{\infty} s_n \omega^{-n}. \tag{10}$$

Then, (7) and (8) become (with $F = \alpha(x)z/2i\omega$)

$$\sum_{n=0}^{\infty}\left(i\omega r_n + c(x)\partial_x r_n - \frac{1}{2}c_x(x)r_n - \frac{\alpha(x)}{2i}(r_{n-1}+s_{n-1})\right)\omega^{-n} = 0,$$
$$\sum_{n=0}^{\infty}\left(i\omega s_n - c(x)\partial_x s_n + \frac{1}{2}c_x(x)s_n - \frac{\alpha(x)}{2i}(r_{n-1}+s_{n-1})\right)\omega^{-n} = 0,$$

where $r_{-1} = s_{-1} = 0$. Defining $v_n = r_n\omega^{-n}$ and $w_n = s_n\omega^{-n}$ and setting each term to zero, we obtain for $x \in (-L,L)$ and $n \geq 0$

$$i\omega v_n + c(x)\partial_x v_n - \frac{1}{2}c_x(x)v_n = -\frac{1}{2i\omega}f_n(x), \tag{11}$$
$$i\omega w_n - c(x)\partial_x w_n + \frac{1}{2}c_x(x)w_n = -\frac{1}{2i\omega}f_n(x), \tag{12}$$

where

$$f_0(x) = \omega f(x), \qquad f_{n+1}(x) = -\alpha(x)(v_n(x)+w_n(x)). \tag{13}$$

We will then approximate $u(x) = v(x) + w(x)$ by $z_m(x)$ obtained by taking the first m terms in the sums in (10),

$$u(x) \approx z_m(x) = \sum_{n=0}^{m}(v_n(x)+w_n(x)). \tag{14}$$

We also need to specify initial conditions for (11) and (12). It follows from Lemma 1 below that z_m will satisfy the boundary conditions (3) and (4) for all m if we let

$$v_n(-L) = \begin{cases} A, & n = 0 \\ 0, & n > 0 \end{cases}, \qquad w_n(L) = 0, \ \ \forall n. \tag{15}$$

To sum up, we solve (11) and (12) for $n = 0,1,2,\ldots,m$ with the given initial conditions (15) and the source function f_n that is defined by (13). Then, we will show (Theorem 1 with Tol $= 0$) that the solution $u(x)$ of the Helmholtz equation (2) can indeed be approximated well by z_m given in (14). Moreover, in contrast to (6), the series (14) converges quickly for large ω.

Remark 1 This is similar to the Bremmer series [7], where

$$i\omega v_n(x) + c(x)\partial_x v_n(x) - \frac{c_x(x)}{2}v_n(x) = -\frac{c_x}{2}w_{n-1}(x),$$
$$i\omega w_n(x) - c(x)\partial_x w_n(x) + \frac{c_x(x)}{2}w_n(x) = \frac{c_x}{2}v_{n-1}(x),$$

with no ω^{-1} in the right hand side. The convergence is more subtle but has been shown in [2,29,5,36]. We prefer (11)–(12) as it clearly separates waves of different size in terms of ω^{-1}, which leads to a simpler analysis.

In a direct implementation, the computational complexity of solving (11,12) would grow algebraically with the frequency ω like for the full Helmholtz equation, since the solution is oscillatory. To get around this we note that (11) and (12) can be simplified when $f_n = 0$. Then, using the ansatz $v_n = Ae^{i\omega\phi}$ in (11) we obtain equations for A and ϕ,

$$\partial_x\phi = \frac{1}{c(x)}, \qquad \partial_x A = \frac{c_x(x)}{2c(x)}A(x). \tag{16}$$

This is in fact GO and can be solved at a cost independent of the frequency. Similar equations can be obtained when the ansatz is used in (12). Thus, the computational cost estimate can be improved by approximating the forcing functions with zero when they are small. More precisely, we do the following. Let $\hat{f}_n$ be the approximate forcing function and $\hat{v}_n$ and $\hat{w}_n$ the corresponding approximate one-way wave equation solutions,

$$i\omega\hat{v}_n + c(x)\partial_x\hat{v}_n - \frac{1}{2}c_x(x)\hat{v}_n = -\frac{1}{2i\omega}\hat{f}_n(x), \tag{17}$$

$$i\omega\hat{w}_n - c(x)\partial_x\hat{w}_n + \frac{1}{2}c_x(x)\hat{w}_n = -\frac{1}{2i\omega}\hat{f}_n(x), \tag{18}$$

with initial data

$$\hat{v}_n(-L) = \begin{cases} A, & n = 0 \\ 0, & n > 0 \end{cases}, \qquad \hat{w}_n(L) = 0, \ \ \forall n. \tag{19}$$

Here, $\hat{f}_n$ is computed as follows. First, calculate the forcing function from the approximate one-way solutions in the same way as before,

$$\tilde{f}_{n+1}(x) = -\alpha(x)(\hat{v}_n(x) + \hat{w}_n(x)). \tag{20}$$

Second, let

$$\mathrm{trunc}(x,\delta) = \begin{cases} 0, & |x| < \delta \\ x, & \text{otherwise} \end{cases}$$

be the truncation function and then define $\hat{f}_n$ as the thresholded version of $\tilde{f}_n$,

$$\hat{f}_n(x) = \mathrm{trunc}(\tilde{f}_n(x), \mathrm{Tol}_n), \qquad \mathrm{Tol}_n = \frac{\omega \mathrm{Tol}}{2^{n+1}L}, \tag{21}$$

for some tolerance Tol. As before, $\tilde{f}_0 = f_0 = \omega f$.

Also for this case we will show (Theorem 1) that the solution $u(x)$ of the Helmholtz equation (2) can be well approximated by

$$u(x) \approx \hat{z}_m(x) := \sum_{n=0}^{m} (\hat{v}_n(x) + \hat{w}_n(x)). \tag{22}$$

Thus, (17)–(18) can be solved independently of the frequency the part of the domain where $\hat{f}_n = 0$. In the part of the domain where $\hat{f}_n \neq 0$ a direct ODE numerical method can be used. The solution of (2) is then approximated by (22). We will also show that the size of the region where $\hat{f}_n \neq 0$ is $\mathscr{O}(1/\omega)$ and that this implies an almost frequency independent computational complexity (Theorem 1).

Hence, the algorithm for computing the solution of (2)–(4) is as follows: Choose some tolerance Tol and, for $n = 0, 1, 2, \ldots$, do the following

1. Replace the function $\tilde{f}_n$ by $\hat{f}_n$ defined by (21). If $\hat{f}_n \equiv 0$ and $n \neq 0$ stop, else
2. Compute $\hat{v}_n$ and $\hat{w}_n$ from (17) and (18). In a domain where $\hat{f}_n \neq 0$ use a direct p-th order numerical method with stepsize

$$\Delta x_f \sim \frac{\mathrm{Tol}^{1/p}}{\omega^{1+1/p}}, \tag{23}$$

otherwise use geometrical optics with stepsize (Figure 1)

$$\Delta x_c \sim \omega \Delta x_f. \tag{24}$$

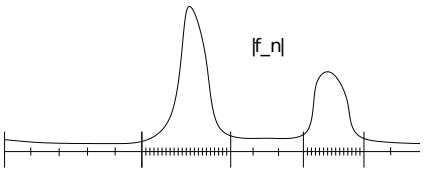

Fig. 1 Function $|f_n|$. In a domain where $|f_n(x)| < \mathrm{Tol}_n$ $\hat{f}_n(x) = 0$ and we can use GO to solve (17) and (18), otherwise $\hat{f}_n(x) = f_n(x)$ and some p-th order ODE numerical scheme can be used.

3. Add $\hat{v}_n + \hat{w}_n$ to $\hat{z}_{n-1}$. Note that in a Runge–Kutta scheme the function $\hat{f}_n$ may need to be evaluated in between the grid points. To obtain those values of $\hat{f}_n$, we use a p-th order interpolation.
4. Compute $\tilde{f}_{n+1}$ from $\hat{v}_n$ and $\hat{w}_n$. Go to 1.

The solution of (2)–(4) is approximated by $\hat{z}_m(x)$.

Remark 2 Note that a direct p-th order method is applied only in intervals of size $\mathscr{O}(\omega^{-1})$. Hence, if Δx_f is given by (23), the computational cost will be $\mathscr{O}(\omega^{1/p})$. See Remark 4 for details.

Remark 3 Extending the method to 2D would involve a number of non-trivial steps. In 2D, waves do not propagate only backward and forward. They propagate asymptotically in the direction $\nabla\phi$ where ϕ solves the eikonal equation (5). A possible generalization of the 1D method could therefore be based on one-way equations of the type

$$
\begin{aligned}
i\omega v_n + c(\mathbf{x})^2 \nabla\phi_n \cdot \nabla v_n - \frac{c(\mathbf{x})^2 \Delta\phi_n}{2} v_n &= \frac{f_n}{2i\omega}, \\
i\omega w_n - c(\mathbf{x})^2 \nabla\phi_n \cdot \nabla w_n - \frac{c(\mathbf{x})^2 \Delta\phi_n}{2} w_n &= \frac{f_n}{2i\omega},
\end{aligned}
$$

with ϕ_n solving (5) and f_n derived in the same way as in 1D, namely

$$
\begin{aligned}
f_{n+1}(\mathbf{x}) &= \alpha(\mathbf{x})(v_n + w_n) + \nabla\phi_n^{\perp} \cdot \nabla \left(c(\mathbf{x})^2 \nabla\phi_n^{\perp} \cdot \nabla(v_n + w_n) \right), \\
\alpha(\mathbf{x}) &= -\nabla \left(\frac{c^2 \Delta\phi_n}{2} \right) \cdot \nabla\phi_n - \frac{c^2 (\Delta\phi_n)^2}{4}.
\end{aligned}
$$

In addition, the boundary conditions for ϕ_{n+1} must be specified based on the solutions v_n and w_n such that $z_n = \sum_{j=0}^{n} (v_j + w_j)$ is a good approximation of the Helmholtz equation, and at the same time the one-way equations can be solved in a computationally cheap way. Another complicating factor is the existence in 2D of caustics, where GO breaks down and ϕ_n should become multi-valued.

3 Analysis

In this section we analyze the method when (17) and (18) are solved exactly. We will first derive estimates of the solution of the equations (11) and (12) and its derivatives. We then show that $\hat{z}_m$ in (22) converges to the solution of (2)–(4). Next, we give an error estimate in terms of m and Tol. As a side effect we obtain a L^∞ estimate of the solution of the Helmholtz

equation in 1D. We also show that the size of the domain where a direct method, which resolves the wavelength, must be used is proportional to $\mathcal{O}(1/\omega)$. More precisely, we prove the following:

Theorem 1 *Let α be defined by (9). Assume $\omega > 1$, $c(x) \in C^2[-L,L]$, and*

$$\frac{\beta|\alpha|_{L_1}}{\omega} \leq \delta_0 < 1, \qquad \beta = \frac{1}{2}\sqrt{\frac{c_{max}}{c_{min}^3}}. \tag{25}$$

If u is the solution of (2)–(4), then

$$|u - \hat{z}_m|_\infty \leq C\left((|A| + |f|_{L_1})\delta_0^{m+1} + \mathrm{Tol}\right). \tag{26}$$

Moreover,

$$\mathrm{meas}\{\hat{f}_n \neq 0\} = \mathrm{meas}\{|\tilde{f}_n| \geq \mathrm{Tol}_n\} \leq \frac{C}{\omega}, \qquad n \geq 1, \tag{27}$$

and

$$|u|_\infty \leq C(|A| + |f|_{L_1}). \tag{28}$$

The constants in (26)–(28) do not depend on ω.

Remark 4 From (27) it follows that for the homogeneous boundary value problem, where $f_0 = \omega f \equiv 0$, the computational cost of our method will not grow fast with ω. More precisely, by (23)–(24),

$$cost \sim \sum_{n=0}^{m+1} \frac{1}{\Delta x_c} + \frac{\mathrm{meas}\{\hat{f}_n \neq 0\}}{\Delta x_f} \sim m\omega^{1/p} = \mathcal{O}(\omega^{1/p}),$$

when $\delta_0^{m+1} \sim$Tol, i.e. $m \sim \log \mathrm{Tol}/\log \delta_0$. More generally, we can allow the forcing function f to depend on ω and be sparse such that

$$\mathrm{meas}\{|f| \geq \mathrm{Tol}\} \leq \frac{C}{\omega}.$$

This covers for instance the physically interesting case of spikes modeling wave sources localized to points, e.g. delta functions or regularized delta functions of type $\omega S(\omega x)$ for a compactly supported $S(y)$.

Remark 5 Note that (25) means that

$$\omega > \frac{1}{2}\sqrt{\frac{c_{max}}{c_{min}^3}}\left|\frac{1}{2}cc_{xx} - \frac{1}{4}c_x^2\right|_{L_1}.$$

This puts a restriction on how fast $c(x)$ can vary relative to the wave length, which is proportional to $1/\omega$. To give a more concrete example of this condition, suppose $c(x)$ changes by $O(1)$ over a small distance ε. Then $c_x \sim 1/\varepsilon$ and $c_{xx} \sim 1/\varepsilon^2$ in an interval of size ε, and $c_x \sim c_{xx} \sim 1$ elsewhere. This would give,

$$|\alpha|_{L_1} = \left|\frac{1}{2}cc_{xx} - \frac{1}{4}c_x^2\right|_{L_1} \sim 1/\varepsilon.$$

Hence, the wave length can be comparable to ε. To compare, recall that in standard geometrical the wave length must be significantly smaller than ε for the approximation to be accurate. In fact, geometrical optics corresponds to one single iteration of the method

($m=0$) and the error estimate above shows that, for this case, $|u-z_0| \sim \delta_0 \sim 1/(\varepsilon\omega)$, while for the general case ($m>0$) with Tol$=0$ we have $|u-z_m| \sim 1/(\varepsilon\omega)^{m+1}$.

Let us also point out that in actual numerical computations there is indeed a cut-off frequency ω_c below which the method does not converge. However, (25) does not give a sharp estimate of ω_c and the method converges also for ω smaller than what is required by (25); see Example 3 in Section 5.

3.1 Estimates of One-Way Solutions

In order to prove Theorem 1, we begin by deriving estimates of the one-way solutions, starting with an estimate for the general linear initial value problem. Consider

$$\frac{dy}{dx} = i\omega a(x,\omega)y + b(x,\omega), \quad x_0 \le x \le x_1, \qquad y(x_0) = y_0, \tag{29}$$

where $a(x,\omega)$ and $b(x,\omega)$ are given functions that depend on the frequency ω. We then have

Theorem 2 *Suppose $\omega > 1$ and that $b(\cdot,\omega) \in C^{p-1}([x_0,x_1])$ for all ω and $a(\cdot,\omega) \in C^{p-1}([x_0,x_1])$ uniformly in $\omega > 1$, i.e. $\sup_{x_0\le x\le x_1} |\partial_x^k a(x,\omega)| < C$, for all $\omega > 1$ and $0 \le k \le p-1$. Moreover, suppose*

$$C_0 = \sup_{\substack{x_0\le t,x\le x_1 \\ \omega>1}} \left| e^{-\omega \int_t^x Im(a(s,\omega))ds} \right| < \infty.$$

Then, if y(x) is a solution of (29),

$$|y|_{L_\infty([x_0,x_1])} \le C_0 \left(|y_0| + |b|_{L_1([x_0,x_1])} \right), \tag{30}$$

$$\left| \partial_x^k y \right|_{L_\infty([x_0,x_1])} \le C_k \left(\omega^k |y_0| + \sum_{i=1}^{k} \omega^{i-1} |\partial_x^{k-i} b(x,\omega)|_{L_\infty([x_0,x_1])} + \omega^k |b|_{L_1([x_0,x_1])} \right),$$

for $k \ge 1$, where C_k are constants that do not depend on ω, $b(x,\omega)$ or y_0.

Proof These estimates are easily obtained from the closed form expression of the exact solution to (29)

$$y(x) = y_0 e^{i\omega \int_{x_0}^x a(s,\omega)ds} + \int_{x_0}^x b(t,\omega) e^{i\omega \int_t^x a(s,\omega)ds} dt.$$

See [41] for details.

We can now derive bounds on the forcing functions f_n and $\tilde{f}_n$ in (13) and (20) as well as the solution and its derivatives of the equations (11) and (12). We prove the following

Theorem 3 *Let $v_n(x)$, $w_n(x)$ be solutions of (11), (12) together with (13), (15) and let β be given by (25). Assume $c(x) \in C^p[-L,L]$ and $\omega > 1$. Then for $p \ge 0$*

$$|\partial_x^p v_n|_\infty \le C\omega^p (|A| + |f|_{L_1}) \left(\frac{\beta}{\omega} |\alpha|_{L_1} \right)^n, \qquad n \ge 0. \tag{31}$$

and

$$|f_n|_{L_1} \le C|\alpha|_{L_1} (|A| + |f|_{L_1}) \left(\frac{\beta}{\omega} |\alpha|_{L_1} \right)^{n-1}, \qquad n \ge 1. \tag{32}$$

Estimate (31) is also valid for $|\partial_x^p w_n|_\infty$. Moreover, if $\tilde{f}_n$ is given by (20) together with (17), (18), (19) then estimate (32) is valid also for $\tilde{f}_n$.

Proof The ODE (11) can be written on the form (29), with $a(x,\omega)$ and $b(x,\omega)$ defined by

$$a(x,\omega) = -\frac{1}{c(x)} - i\frac{c_x(x)}{2c(x)\omega}, \qquad b(x,\omega) = -\frac{1}{2i\omega c(x)} f_n(x),$$

which satisfy the assumptions in Theorem 2 with $x_0 = -L$, $x_1 = L$ and

$$C_0 = \sup_{-L\le t,x\le L} \left| e^{\int_t^x \frac{c_s}{2c(s)} ds} \right| = \sup_{-L\le t,x\le L} \left| e^{\int_t^x d\left(\ln(\sqrt{c(s)})\right) ds} \right| = \sup_{-L\le t,x\le L} \frac{\sqrt{c(x)}}{\sqrt{c(t)}} \le 2\beta c_{\min},$$

where β is the constant defined by (25). For $|b|_{L_1}$, we can estimate

$$|b|_{L_1} = \left| -\frac{1}{2i\omega c(x)} f_n(x) \right|_{L_1} \le \frac{1}{2\omega c_{\min}} |f_n|_{L_1}.$$

Hence, from Theorem 2 with $p = 0$,

$$|v_n|_\infty \le 2\beta c_{\min} \left(|v_n(-L)| + \frac{1}{2\omega c_{\min}} |f_n|_{L_1} \right) \le C|v_n(-L)| + \frac{\beta}{\omega} |f_n|_{L_1}.$$

Moreover, $f_0(x) = \omega f(x)$ and $v_n(-L)$ is given by (15). Thus,

$$|v_0|_\infty \le C|A| + \beta |f|_{L_1} \le C(|A| + |f|_{L_1}), \qquad |v_n|_\infty \le \frac{\beta}{\omega} |f_n|_{L_1}, \quad n \ge 1. \tag{33}$$

We obtain the corresponding results for w_n in the same way. Using (13), we can estimate

$$|f_{n+1}|_{L_1} = \int_{\mathbb{R}} |\alpha(x)|\, |v_n(x) + w_n(x)|\, dx \le |\alpha|_{L_1} (|v_n|_\infty + |w_n|_\infty). \tag{34}$$

Then, for $n = 1$

$$|f_1|_{L_1} \le |\alpha|_{L_1} (|v_0|_\infty + |w_0|_\infty) \overset{(33)}{\le} C|\alpha|_{L_1} (|A| + |f|_{L_1})$$

and we use induction to prove that (32) is also valid for $n \ge 2$. So, assume that (32) is true for $n = k$ and show that it is true for $n = k+1$,

$$|f_{k+1}|_{L_1} \le |\alpha|_{L_1} (|v_k|_\infty + |w_k|_\infty) \overset{(33)}{\le} \frac{\beta}{\omega} |f_k|_{L_1} |\alpha|_{L_1} \le C|\alpha|_{L_1} (|A| + |f|_{L_1}) \left(\frac{\beta}{\omega} |\alpha|_{L_1} \right)^k.$$

Hence, (32) is proved. We now note that the only property connecting f_n with v_n, w_n and f that has been used so far is (34) and the fact that $|f_0|_{L_1} \le \omega |f|_{L_1}$. The same relationships clearly hold also for $\hat{f}_n$, $\hat{v}_n$, $\hat{w}_n$ and f. Therefore (32) holds also for $\tilde{f}_n$.

We continue with proving (31). For $p = 0$ it follows from (33) and (32). For $p \ge 1$, we again use mathematical induction and Theorem 2. Assuming (31) is true for derivatives up to the order p we show that it is also true for the derivative of order $p+1$. From Theorem 2, it follows

$$|\partial_x^{p+1} v_n|_\infty \le C_{p+1} \left(\omega^{p+1} |v_n(-L)| + \sum_{k=1}^{p+1} \omega^{k-1} \left| \partial_x^{p+1-k} \frac{f_n}{2i\omega c} \right|_\infty + \omega^{p+1} \left| \frac{f_n}{2i\omega c} \right|_{L_1} \right). \tag{35}$$

Before we continue, let us estimate

$$\begin{aligned}\sum_{k=1}^{p+1}\omega^{k-1}\left|\partial_x^{p+1-k}\left(\frac{f_n}{2i\omega c}\right)\right|_\infty &= \sum_{k=0}^{p}\omega^k\left|\partial_x^{p-k}\left(\frac{\alpha}{2i\omega c}(v_{n-1}+w_{n-1})\right)\right|_\infty \\ &= \sum_{k=0}^{p}\omega^k\left|\sum_{m=0}^{p-k}\binom{p-k}{m}\partial_x^m(v_{n-1}+w_{n-1})\partial_x^{p-k-m}\left(\frac{\alpha}{2i\omega c}\right)\right|_\infty \\ &\le \frac{C}{\omega}(|A|+|f|_{L_1})\left(\frac{\beta}{\omega}|\alpha|_{L_1}\right)^{n-1}\sum_{k=0}^{p}\omega^k\sum_{m=0}^{p-k}\omega^m \le C(|A|+|f|_{L_1})\left(\frac{\beta}{\omega}|\alpha|_{L_1}\right)^n\omega^p\end{aligned}$$

and, with $\delta=\beta|\alpha|_{L_1}/\omega$,

$$\left|\frac{f_n}{2i\omega c}\right|_{L_1} \le \frac{C}{\omega}|\alpha|_{L_1}(|A|+|f|_{L_1})\delta^{n-1} \le C(|A|+|f|_{L_1})\delta^n.$$

Thus if $\lambda_0=1$ and $\lambda_n=0$ for $n>0$, we get from (35)

$$|\partial_x^{p+1}v_n|_\infty \le C_{p+1}\left(\omega^{p+1}|A|\lambda_n+\omega^{p+1}(|A|+|f|_{L_1})\delta^n\right) \le C\omega^{p+1}(|A|+|f|_{L_1})\delta^n,$$

which is what we wanted to show.

3.2 Convergence of the Algorithm

In this section we show two lemmas about the convergence of the algorithm. These are subsequently used in the proof of Theorem 1.

Lemma 1 *Let $f(x)\in C^1[-L,L]$ and $c(x)\in C^2[-L,L]$. If $v(x)$ and $w(x)$ satisfy (11) and (12) with boundary conditions $v(-L)=A$ and $w(L)=0$, then $z(x)=v(x)+w(x)$ satisfies*

$$c^2(x)z_{xx}(x)+\omega^2 z(x)=f(x)+\alpha(x)(v(x)+w(x))$$

with boundary conditions:

$$\begin{aligned}c(-L)z_x(-L)-i\omega z(-L)&=-2i\omega A,\\ c(L)z_x(L)+i\omega z(L)&=0.\end{aligned}$$

Proof This follows from simple mathematical manipulations, see [41] for details.

Lemma 2 *If $\frac{\beta|\alpha|_{L_1}}{\omega}\le\delta_0<1$, the sequence $\{z_m\}_{m=1}^\infty$ defined by (14) with $c(x)\in C^2[-L,L]$ converges in $C^2[-L,L]$ when $m\to\infty$ and its limit z satisfies the Helmholtz equation (2) with boundary conditions (3) – (4). Moreover,*

$$|z|_\infty \le C(|A|+|f|_{L_1}), \tag{36}$$

where C is a constant that depends on δ_0 and $c(x)$ but not on ω.

Proof Let v_j, w_j and f_j be given by (11), (12) together with (13), (15). From Lemma 1 it follows that $v_j + w_j$ satisfies

$$c^2(v_j+w_j)_{xx} + \omega^2(v_j+w_j) = f_j - f_{j+1}, \quad j = 0,\dots,m \tag{37}$$

with boundary conditions

$$c(-L)(v_j(-L)+w_j(-L))_x - i\omega(v_j(-L)+w_j(-L)) = -2i\omega v_j(-L), \tag{38}$$

$$c(L)(v_j(L)+w_j(L))_x + i\omega(v_j(L)+w_j(L)) = 0. \tag{39}$$

Summing the equations (37) for $j = 0,\dots,m$, we get that z_m satisfies the following equation:

$$c^2\partial_{xx}z_m + \omega^2 z_m = f_0 - f_{m+1}.$$

Moreover, summing (38) and (39), we get that z_m satisfies the boundary conditions:

$$c(-L)\partial_x z_m(-L) - i\omega z_m(-L) = -2i\omega A, \tag{40}$$

$$c(L)\partial_x z_m(L) + i\omega z_m(L) = 0. \tag{41}$$

Using Theorem 3, we can show that for $p = 0,1,2$,

$$|\partial_x^p z_m - \partial_x^p z_n|_\infty \le \sum_{j=min(n,m)}^{max(n,m)} |\partial_x^p v_j + \partial_x^p w_j|_\infty \le \sum_{j=min(n,m)}^{max(n,m)} C\omega^p(|A|+|f|_{L_1})\left(\frac{\beta}{\omega}|\alpha|_{L_1}\right)^j$$
$$\le C\omega^p \delta_0^{min(n,m)} \to 0, \quad m,n \to \infty.$$

This means that $\{z_m\}_{m=0}^\infty$ is a Cauchy sequence in $\mathrm{C}^2[-L,L]$ and there exists $z \in \mathrm{C}^2[-L,L]$ such that $z_m \to z$ as $m \to \infty$. Moreover, since $|f_{m+1}|_\infty \le C(|v_m|_\infty + |w_m|_\infty) \to 0$ as $m \to \infty$ by Theorem 3, it follows that z satisfies the Helmholtz equation (2),

$$c^2 z_{xx} + \omega^2 z = \omega f,$$

and by taking $\lim_{m\to\infty}$ of (40) and (41), it follows that z satisfies the boundary conditions (3),(4). Finally, by Theorem 3,

$$|z_m|_\infty \le \sum_{j=0}^m (|v_j|_\infty + |w_j|_\infty) \le C(|A|+|f|_{L_1}) \sum_{j=0}^m \left(\frac{\beta}{\omega}|\alpha|_{L_1}\right)^j \le C(|A|+|f|_{L_1}) \sum_{j=0}^m \delta_0^j.$$

By taking the limit $m \to \infty$ in both sides of this inequality we obtain (36).

3.3 Proof of Theorem 1

We can now prove Theorem 1. Let u be the solution of (2) with boundary conditions (3)-(4). If $\hat{v}_m$ and $\hat{w}_m$ are solutions of (17) and (18) then from Lemma 1 it follows that $\hat{v}_m(x) + \hat{w}_m(x)$ satisfies

$$c^2(\hat{v}_m+\hat{w}_m)_{xx} + \omega^2(\hat{v}_m+\hat{w}_m) = \hat{f}_m - \tilde{f}_{m+1},$$

with boundary conditions

$$c(-L)(\hat{v}_m(-L)+\hat{w}_m(-L))_x - i\omega(\hat{v}_m(-L)+\hat{w}_m(-L)) = -2i\omega\hat{v}_m(-L),$$

$$c(L)(\hat{v}_m(L)+\hat{w}_m(L))_x + i\omega(\hat{v}_m(L)+\hat{w}_m(L)) = 0.$$

Let e_m and $\hat{e}_m$ solve

$$c^2\partial_{xx}e_m+\omega^2 e_m=\hat{f}_m,\qquad c^2\partial_{xx}\hat{e}_m+\omega^2\hat{e}_m=\tilde{f}_m-\hat{f}_m$$

with boundary conditions (3)-(4) with $A=0$. By the uniqueness of the solution of the Helmholtz equation with boundary conditions (3)-(4) and the fact that $\tilde{f}_0=f_0$, it follows that $u=e_0+\hat{e}_0$ and

$$e_m=\hat{v}_m+\hat{w}_m+e_{m+1}+\hat{e}_{m+1}.$$

Hence, by induction,

$$u=e_0+\hat{e}_0=\hat{v}_0+\hat{w}_0+...+\hat{v}_m+\hat{w}_m+e_{m+1}+\sum_{j=0}^{m+1}\hat{e}_j=\hat{z}_m+e_{m+1}+\sum_{j=0}^{m+1}\hat{e}_j.$$

Thus, if $\hat{z}_m$ is given by (22),

$$|u-\hat{z}_m|_\infty<|e_{m+1}|_\infty+\sum_{j=0}^{m+1}|\hat{e}_j|_\infty. \tag{42}$$

Let us estimate $|e_m|_\infty$. Since e_m satisfies the Helmholtz equation, from Lemma 2 and Theorem 3 it follows

$$|e_m|_\infty\le C\left(|A|+\frac{1}{\omega}|\hat{f}_m|_{L_1}\right)\overset{A=0}{\le}\frac{C}{\omega}|\hat{f}_m|_{L_1}\le\frac{C}{\omega}|\tilde{f}_m|_{L_1}\le C\left(|A|+|f|_{L_1}\right)\left(\frac{\beta}{\omega}|\alpha|_{L_1}\right)^m. \tag{43}$$

Let us now estimate $|\hat{e}_m|_\infty$. Again, using Lemma 2, we conclude

$$|\hat{e}_m|_\infty\le C\left(|A|+\frac{1}{\omega}|\tilde{f}_m-\hat{f}_m|_{L_1}\right)\overset{A=0,m\ge1}{=}\frac{C}{\omega}|\tilde{f}_m-\hat{f}_m|_{L_1}\le\frac{C}{\omega}\mathrm{Tol}_m. \tag{44}$$

Now, using (43), (44) and (21), (42) becomes

$$|u-\hat{z}_m|_\infty\le C\left(|A|+|f|_{L_1}\right)\left(\frac{\beta}{\omega}|\alpha|_{L_1}\right)^{m+1}+C\sum_{j=0}^{m+1}\frac{\mathrm{Tol}}{2^{j+1}}\le C((|A|+|f|_{L_1})\delta_0^{m+1}+\mathrm{Tol})$$

which we wanted to show.

Let us finally show (27). Define,

$$g(y)=\mathrm{meas}\{x\in[-L,L]:|\hat{f}_n(x)|\ge y\}.$$

Clearly, $g(\mathrm{Tol}_n)\le g(0)=2L$ and $g(y)=0$ for $y\ge|\tilde{f}_n|_\infty$, while for $0<y<|\tilde{f}_n|_\infty$ it is a decreasing function. Then,

$$yg(y)\le 2L|\tilde{f}_n|_\infty\le 2LC|\alpha|_\infty(|A|+|f|_{L_1})\left(\frac{\beta}{\omega}|\alpha|_{L_1}\right)^{n-1}\le C\left(\frac{\beta}{\omega}|\alpha|_{L_1}\right)^{n-1}$$

and thus, using (21),

$$g(\mathrm{Tol}_n)\le\frac{C}{\mathrm{Tol}_n}\left(\frac{\beta}{\omega}|\alpha|_{L_1}\right)^{n-1}\le\frac{C}{\omega},$$

which proves (27). The estimate (28) follows directly from Lemma 2. Thus, Theorem 1 is proved.

4 Error Analysis for Numerical Implementation

In this section we derive the error of the method when also the numerical approximation is taken into account. Our main result is Theorem 4 where we show that the error depends only weakly on the frequency for a fixed computational cost.

In Section 3, we showed that the solution of (2)–(4) can be approximated by the sum z_m defined by (14). Now, we want to solve for $z_m(x)$ numerically. That means that we have to solve equations (11)–(12). As explained in Section 2 we solve them directly when $|f_n| > \mathrm{Tol}_n$. The equations are then of the form (29). For (11), $a(x,\omega)$ and $b(x,\omega)$ are defined by

$$a(x,\omega) = -\frac{1}{c(x)} - i\frac{c_x(x)}{2c(x)\omega}, \qquad b^n(x,\omega) = -\frac{1}{2i\omega c(x)} f^n(x). \tag{45}$$

If $|f_n| < \mathrm{Tol}_n$, then (11)–(12) are transformed into (16). In both cases we use a p-th order Runge–Kutta (R-K) scheme; for (11)–(12) we use a fine mesh size Δx_f while for (16) we use a coarse mesh size Δx_c. Note that in (45) we use $b^n(x,\omega)$ and not $b(x,\omega)$ because the function depends on $f^n(x)$ which depends on n.

4.1 Numerical Implementation

Before we describe the details of how the equations are solved numerically, let us introduce some notation. We denote the grid in iteration n by $G^n = \{x_1^n, x_2^n, \dots\}$. Note that the grid will change in each iteration and thus depends on n. We further introduce grid functions that approximate the corresponding exact functions in the grid points, the one-way solutions $v_j^n \approx v^n(x_j^n)$, $w_j^n \approx w^n(x_j^n)$ and the amplitude/phase solutions in (16), $A_j^n \approx A^n(x_j)$, $\phi_j^n \approx \phi^n(x_j)$.

When we change grids between iterations we need to interpolate the results from the previous grid onto the next one. We let $\bar{v}^n(x)$ be an interpolant of the numerically computed sequence $\{v_j^n\}$ on G^n. Similarly, we define $\bar{w}^n(x)$ as the interpolant of $\{w_j^n\}$. Furthermore, we set

$$\bar{f}^{n+1}(x) := -\alpha(x)\left(\bar{v}^n(x) + \bar{w}^n(x)\right), \qquad \bar{b}^n(x,\omega) = -\frac{1}{2i\omega c(x)}\bar{f}^n(x) \tag{46}$$

and

$$\bar{z}^m(x) = \sum_{n=0}^{m} (\bar{v}^n(x) + \bar{w}^n(x)). \tag{47}$$

Note that since $f(x)$ in the equation is known we do not need to interpolate in the first step and thus have

$$f^0(x) = \bar{f}^0(x), \qquad b^0(x) = \bar{b}^0(x). \tag{48}$$

4.1.1 Construction of the Grid

The full grid G^n in iteration n will have one coarse part, G_c^n, where the mesh size is Δx_c and one fine part G_f^n with the mesh size Δx_f. We define the reference fine and coarse grids, G_f^{ref} and G_c^{ref},

$$G_f^{\mathrm{ref}} = \{x_j^f = -L + j\Delta x_f, j = 0, \dots J_f\}, \qquad \Delta x_f = \frac{2L}{J_f},$$

$$G_c^{\mathrm{ref}} = \{x_j^c = -L + j\Delta x_c, j = 0, \dots J_c\}, \qquad \Delta x_c = \frac{2L}{J_c}.$$

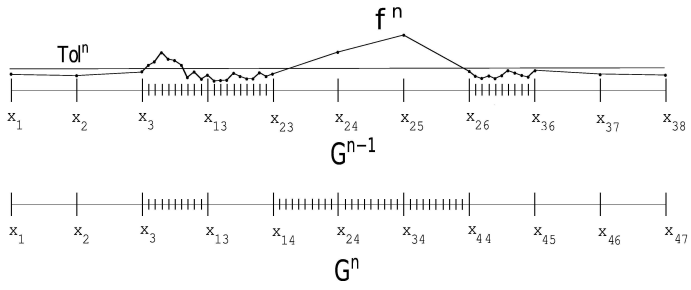

Fig. 2 Grids G^{n-1} (top) and G^n (bottom). Here, $G^n = G_c^n \cup G_f^n$ where $G_c^n = \{x_1, x_2, x_3, x_{13}, x_{14}, x_{44}, \ldots, x_{47}\}$ and $G_f^n = \{x_3, \ldots x_{13}, x_{14}, \ldots, x_{44}\}$.

For simplicity, we assume $\Delta x_c/\Delta x_f \in \mathbb{Z}$, so that every coarse point is also a fine point, i.e. $G_c^{\text{ref}} \subset G_f^{\text{ref}}$. Initially we take $G_c^0 = G_c^{\text{ref}}$ and $G_f^0 = \emptyset$, since $f^0 \equiv 0$. Assume now that we have G^{n-1} and the approximate solutions v_j^{n-1} and w_j^{n-1}. Let

$$F_j = |\alpha(x_j)| \left(|v_j^{n-1}| + |w_j^{n-1}| \right),$$

which corresponds to the f^n computed on the grid G^{n-1}. Then G^n is defined in the following way. First, let R^n be the indices of the coarse intervals that contain at least one "large" F_j value,

$$R^n = \{m : \exists x_j^{n-1} \in [x_m^c, x_{m+1}^c],\ F_j > \text{Tol}^n\}.$$

The grid is then constructed by taking a fine grid in the intervals given by R^n and a coarse grid elsewhere, (Figure 2)

$$G_f^n = G_f^{\text{ref}} \cap \{x \in [x_m^c, x_{m+1}^c] : m \in R^n\}, \qquad G_c^n = G_c^{\text{ref}} \cap \{x \in [x_m^c, x_{m+1}^c] : m \notin R^n\}.$$

Finally, $G^n = G_f^n \cup G_c^n$.

4.1.2 Solving One-Way Equations on the Fine Grid

Let us now explain the numerical approximation of (11) for $x_j \in G_f^n$. For simplicity, we suppress henceforth the iteration index on the grid points, i.e. we use x_j instead of x_j^n. With $a(x,\omega)$ and $b^n(x,\omega)$ defined above, (17) becomes

$$\partial_x v^n(x) = i\omega a(x,\omega) v^n(x) + b^n(x,\omega). \tag{49}$$

We solve (49) with a p-th order explicit s-stage R-K scheme, where $0 \le p \le s$. From Theorem 5 in Section 4.3.1 it follows that, whenever $b^n(x,\omega)$ is available explicitly, the R-K scheme applied to our problem reads

$$v_{j+1}^n = (1 + i\Delta x_f \omega \tilde{a}(x_j, \Delta x_f, \omega)) v_j^n + \tilde{b}^n(x_j, \Delta x_f, \omega), \qquad x_j \in G_f^n$$

where $\tilde{a}(x, \Delta x_f, \omega)$ and $\tilde{b}(x, \Delta x_f, \omega)$ are some functions that depend on the R-K method used. For example,

$$\tilde{a}(x, \Delta x_f, \omega) = \frac{1}{2}\left(a(x,\omega) + a(x + \Delta x_f, \omega) + i\omega a(x,\omega) a(x + \Delta x_f, \omega) \right),$$
$$\tilde{b}^n(x, \Delta x_f, \omega) = \frac{1}{2}\left((1 + i\omega a(x + \Delta x_f, \omega)) b^n(x,\omega) + b^n(x + \Delta x_f, \omega) \right),$$

when (49) is solved with the second order R-K method. Since $b^n(x,\omega)$ is only known explicitly for $n=0$, we use

$$v_{j+1}^n = (1+i\Delta x_f \omega \tilde{a}(x_j,\Delta x_f,\omega))v_j^n + \tilde{\bar{b}}^n(x_j,\Delta x_f,\omega), \qquad n \geq 1,$$

where $\tilde{\bar{b}}^n(x,\Delta x_f,\omega)$ is computed as follows. First, compute $\bar{b}^n(x,\omega)$ by (46). Second, compute $\tilde{\bar{b}}^n(x,\Delta x_f,\omega)$ from $\bar{b}^n(x,\omega)$ in the same way as $\tilde{b}^n(x,\Delta x_f,\omega)$ is computed from $b^n(x,\omega)$. For example,

$$\tilde{\bar{b}}^n(x,\Delta x_f,\omega) := \frac{1}{2}\left((1+i\omega a(x+\Delta x_f,\omega))\bar{b}^n(x,\omega) + \bar{b}^n(x+\Delta x_f,\omega)\right)$$

in the case of the second order R-K method. The solution of (12) is computed in a similar way.

4.1.3 Solving One-Way Equations on the Coarse Grid

Let us now explain the numerical implementation of GO equations (16) which are used for $x_j \in G_c^n$. Discretizing the equations with the R-K method, we obtain

$$\phi_{j+1}^n = \phi_j^n + \Delta x_c \breve{\bar{b}}^n(x_j,\Delta x_c), \qquad A_{j+1}^n = (1+\Delta x_c \breve{a}(x_j,\Delta x_c))A_j^n, \qquad x_{j+1}-x_j = \Delta x_c,$$

where $\breve{a}(x_j,\Delta x_c)$ and $\breve{\bar{b}}^n(x_j,\Delta x_c)$ are functions obtained in the same way as $\tilde{a}(x_j,\Delta x_f,\omega)$ and $\tilde{\bar{b}}^n(x_j,\Delta x_f,\omega)$. The solution of (17) for $\hat{f}_n = 0$ is then computed by

$$v_{j+1}^n = A_{j+1}^n e^{i\omega\phi_{j+1}^n} = e^{i\omega\Delta x_c \breve{\bar{b}}^n(x_j,\Delta x_c)}(1+\Delta x_c \breve{a}(x_j,\Delta x_c))v_j^n, \qquad x_{j+1}-x_j = \Delta x_c.$$

The solution of (12) for $x_j \in G_c^n$ can be obtained in a similar way.

4.1.4 Final Formulation

If we let J^n be the total number of discretization points in G^n we can define the one-step solution operator

$$\mathscr{S}^n(x_j,\omega) = \begin{cases} 1+i\Delta x_f \omega \tilde{a}(x_j,\Delta x_f,\omega), & x_{j+1}-x_j = \Delta x_f \\ (1+\Delta x_c \breve{a}(x_j,\Delta x_c))e^{i\omega\Delta x_c \breve{\bar{b}}^n(x_j,\Delta x_c)}, & x_{j+1}-x_j = \Delta x_c. \end{cases} \tag{50}$$

The solution of (11) is then computed by

$$v_{j+1}^n = \mathscr{S}^n(x_j,\omega)v_j^n + \lambda_j^n \Delta x_f \tilde{\bar{b}}^n(x_j,\Delta x_f,\omega), \qquad j=0,\ldots,J^n-1$$

where

$$\lambda_j^n = \begin{cases} 1, & x_{j+1}-x_j = \Delta x_f \\ 0, & x_{j+1}-x_j = \Delta x_c. \end{cases}$$

Remark 6 If the grid G^n is given as in Figure 2, we solve for v^n in the following way:

1. Solve GO equations in $[x_1,x_3]$,
2. Use v_3^n as initial condition and solve (11) in $[x_3,x_{13}]$
3. Use v_{13}^n as initial condition and solve GO equations in $[x_{13},x_{14}]$
4. Use v_{14}^n as initial condition and solve (11) in $[x_{14},x_{44}]$
5. Use v_{44}^n as initial condition and solve GO equations in $[x_{44},x_{47}]$.

4.2 The Main Theorem

The main result of this section is the following theorem about the error in the numerical approximation computed with the method described in Section 4.1, more precisely the error in the interpolant of the numerical result $\bar{z}_m(x)$ in (47).

Theorem 4 *Assume $\Delta x_f^p \omega^{p+1} < 1$ and as in the continuous case,*

$$\frac{\beta|\alpha|_{L_1}}{\omega} \leq \delta_0 < 1. \tag{51}$$

Moreover, we assume

$$\omega > C_5 + C_6 C_7, \tag{52}$$

where these constants are defined in Lemma 3, Lemma 4 and Lemma 5 (see below). Let u be the solution of equation (2). Then, there exists a $\delta < 1$ and a constant C such that

$$|u(x) - \bar{z}^m(x)|_\infty \leq C\left((|A| + |f|_{L_1})\delta^{m+1} + \Delta x_f^p \omega^{p+1} + \Delta x_c^p \omega + \mathrm{Tol}\right). \tag{53}$$

Remark 7 If we take

$$\Delta x_f = \frac{\mathrm{Tol}^{1/p}}{\omega^{1+1/p}}, \qquad \Delta x_c = \omega \Delta x_f,$$

we get from (53)

$$|u(x) - \bar{z}^m(x)|_\infty \leq C((|A| + |f|_{L_1})\delta^{m+1} + \mathrm{Tol})$$

and as in Remark 4 we need a fixed ω-independent number of iterations m to reduce the first term to be less than Tol.

4.3 Error Analysis

In this section, we first introduce Runge–Kutta methods and an error estimate. Then we derive error estimates for the numerical solution of the one-way wave equations. Throughout this section, we assume

$$\Delta x_f < \frac{1}{\omega^{1+1/p}} \quad (i.e.\ \Delta x_f^p \omega^{p+1} < 1) \tag{54}$$

and $c(x) \in C^p[-L, L]$ and $\omega > 1$. In particular, this means that $\Delta x_f \omega < 1$.

4.3.1 Runge–Kutta Schemes

We start with a general result on R-K schemes applied to linear ordinary equations of the type

$$y' = i\omega a(x,\omega)y + b(x,\omega), \qquad y(0) = y_0, \tag{55}$$

where we assume that $b(\cdot,\omega) \in C^{p+1}(\mathbb{R})$ for all ω and $a(\cdot,\omega) \in C^{p+1}(\mathbb{R})$ uniformly in ω, i.e. that

$$\sup_{x,\omega\in\mathbb{R}} |\partial_x^k a(x,\omega)| \leq C, \qquad k = 0, \ldots, p+1.$$

This means in particular that the solution $y(x) \in C^{p+2}$. We consider a p-th order explicit s-stage R-K scheme, where $0 \le p \le s$, described by

$$\xi_1 = i\omega a(x_j, \omega) y_j + b(x_j, \omega),$$

and for $k = 2, \ldots, s$,

$$\xi_k = i\omega a(x_j + \gamma_k \Delta x, \omega) \left(y_j + \Delta x \sum_{\ell=1}^{k-1} \alpha_{k,\ell} \xi_\ell \right) + b(x_j + \gamma_k \Delta x, \omega).$$

Finally,

$$y_{j+1} = y_j + \Delta x \sum_{k=1}^{s} \beta_k \xi_k,$$

where $\alpha_{k,l}, \gamma_k, \beta_k \in \mathbb{R}$ and $\gamma_k = \sum_l \alpha_{k,l}$, $0 \le \gamma_k \le 1$. The error in one step, the local truncation error is then defined by

$$\tau_j := y(x_j + \Delta x) - y_{j+1},$$

when $y(x)$ is the exact solution of (55) and $y(x_j) = y_j$. The following theorem is given without proof. The proof can be found in [41].

Theorem 5 *Suppose $\Delta x \omega \le 1$ and $\omega \ge 1$. Then the local truncation error can be estimated as*

$$|\tau_j| \le C\, \Delta x^{p+1} \sum_{\ell=0}^{p+1} |y^{(p+1-\ell)}|_{L^\infty[x_n, x_{n+1}]} \omega^\ell,$$

where the constant is independent of x, ω, $y(x)$ and $b(x, \omega)$. Moreover, if $|\partial_x^p y| \le B(\omega)\omega^p$, then

$$|\tau_j| \le CB(\omega)(\Delta x \omega)^{p+1}.$$

Moreover, the Runge–Kutta scheme can be written in the following form

$$y_{j+1} = y_j + \Delta x \left[i\omega \tilde{a}(x_j, \omega, \Delta x) y_j + \sum_{k=1}^{s} \tilde{r}_k(x_j, \omega, \Delta x) b(x_j + \gamma_k \Delta x, \omega) \right],$$

where $\tilde{a}$ and $\tilde{r}_k$ do not depend on y_j or $b(x, \omega)$ and $\tilde{r}_k(x, \omega, \Delta x)$ is bounded in ω and Δx as long as $\Delta x \omega \le 1$.

4.3.2 Error Estimate for One-Way Equations

Now we derive an error estimate for the one-way equations. Let us first introduce some notation. Let τ_j^n be the local truncation error at x_j in iteration n, i.e.,

$$\tau_j^n = v^n(x_{j+1}) - v_{j+1}^n,$$

given that for the exact solution $v^n(x_j) = v_j^n$ and let τ_f^n and τ_c^n be the max local truncation errors for the fine and the coarse grid respectively, i.e.,

$$\tau_f^n = \max_{x_j \in G_f^n} |\tau_j^n|, \qquad \tau_c^n = \max_{x_j \in G_c^n} |\tau_j^n|.$$

Let $\varepsilon_{v(j)}^n$ (and ε_v^n) be the (maximum) global numerical error of v_j^n, i.e.

$$\varepsilon_{v(j)}^n = |v^n(x_j) - v_j^n| \quad j = 0, \ldots, J^n, \qquad \varepsilon_v^n = \max_j \varepsilon_{v(j)}^n, \tag{56}$$

where J^n is the number of discretization points at iteration n. In the same way we define $\varepsilon^n_{w(j)}$ and ε^n_w. Finally, let

$$\varepsilon^n = \varepsilon^n_v + \varepsilon^n_w.$$

For the interpolation we use a p-th order method and we make the following two assumptions.

– (A1) The interpolated values are bounded by the max error in the point values and an interpolation error governed by the smoothness of the exact solution,

$$\begin{aligned} |v^n(x) - \bar{v}^n(x)| &\le C\left(\varepsilon^n_v + \max(\Delta x_f^p |\partial_x^p v^n|_\infty, \omega \Delta x_c^p |v_n|_\infty)\right) \\ &\le C\left(\varepsilon^n_v + \max((\omega\Delta x_f)^p, \omega\Delta x_c^p)\delta_0^n\right), \end{aligned} \tag{57}$$

where the second step follows from Theorem 3. The same estimate is assumed to hold for $w^n - \bar{w}^n$. As a consequence, when $n \ge 1$,

$$|f^n(x) - \bar{f}^n(x)| \le C\alpha(x)\left(\varepsilon^{n-1} + \max((\omega\Delta x_f)^p, \omega\Delta x_c^p)\delta_0^{n-1}\right). \tag{58}$$

Again, note that $f^0(x) = \bar{f}^0(x)$ by (48).

– (A2) The interpolated value $\bar{f}^n(x)$ is bounded by Tol_n when x is in the coarse grid part,

$$\max_{x_\ell \le x \le x_{\ell+1}} |\bar{f}^n(x)| \le C\mathrm{Tol}_n, \qquad x_\ell, x_{\ell+1} \in G_c^n. \tag{59}$$

We are now ready to analyze the local truncation errors.

Lemma 3 *The max local truncation error can be estimated*

$$\tau_f^n \le C\left(\Delta x_f \omega\right)^{p+1} \delta_0^n, \tag{60}$$

$$\tau_c^n \le C_5 \Delta x_c \begin{cases} \max\left((\Delta x_f\omega)^p, \omega\Delta x_c^p\right)\delta_0^n + \frac{\varepsilon^{n-1}+\mathrm{Tol}_n}{\omega}, & n \ge 1, \\ \omega\Delta x_c^p \delta_0^n + \frac{\mathrm{Tol}_n}{\omega}, & n = 0, \end{cases} \tag{61}$$

where C and C_5 are some constants that do not depend on ω.

Proof The estimate (60) follows directly from Theorem 3 and Theorem 5. Let us now prove (61). Let $\tilde{v}^n$ be the exact solution of (11) for $f^n = 0$, i.e. the solution of the equation that is solved on the coarse grid G_c^n. Assume $v^n(x_j) = \tilde{v}^n(x_j) = v_j^n$. Then,

$$\begin{aligned} \tau_j^n &= |v^n(x_{j+1}) - v_{j+1}^n| = |v^n(x_{j+1}) - \mathscr{S}^n(x_j,\omega)v^n(x_j)| \\ &\le |v^n(x_{j+1}) - \tilde{v}^n(x_{j+1})| + |\tilde{v}^n(x_{j+1}) - \mathscr{S}^n(x_j,\omega)v^n(x_j)|. \end{aligned}$$

We note that $v^n - \tilde{v}^n$ satisfies the same equation as v^n on $[x_j, x_{j+1}]$ with $v^n(x_j) - \tilde{v}^n(x_j) = 0$, so from (30), it follows that for $n \ge 1$,

$$|v^n(x_{j+1}) - \tilde{v}(x_{j+1})| \le \frac{\beta}{\omega}|f^n|_{L_1([x_j,x_{j+1}])} \le \frac{\beta}{\omega}(|f^n - \bar{f}^n|_{L_1([x_j,x_{j+1}])} + |\bar{f}^n|_{L_1([x_j,x_{j+1}])}). \tag{62}$$

From (59) we have

$$\frac{\beta}{\omega}|\bar{f}^n|_{L_1([x_j,x_{j+1}])} \le \frac{\beta\Delta x_c}{\omega}|\bar{f}^n|_{L_\infty([x_j,x_{j+1}])} \le \frac{C\beta\Delta x_c}{\omega}\mathrm{Tol}_n, \tag{63}$$

and by (58),

$$\frac{\beta}{\omega}|f^n-\bar{f}^n|_{L_1([x_j,x_{j+1}])}\le\begin{cases}C\frac{\Delta x_c\beta|\alpha|_{L_1}}{\omega}\left[\max\left((\Delta x_f\omega)^p,\omega\Delta x_c^p\right)\delta_0^{n-1}+\varepsilon^{n-1}\right], & n\ge 1,\\ 0, & n=0,\end{cases}$$
$$=\begin{cases}C\Delta x_c\left[\max\left((\Delta x_f\omega)^p,\omega\Delta x_c^p\right)\delta_0^{n}+\frac{C_\tau}{\omega}\varepsilon^{n-1}\right], & n\ge 1,\\ 0, & n=0.\end{cases}\tag{64}$$

For $|\tilde{v}^n(x_{j+1})-\mathscr{S}^n(x_j,\omega)v^n(x_j)|$ we calculate

$$\begin{aligned}|\tilde{v}^n(x_{j+1})-\mathscr{S}^n(x_j,\omega)v^n(x_j)| &= \left|A^n(x_{j+1})e^{i\omega\phi^n(x_{j+1})}-A^n_{j+1}e^{i\omega\phi^n_{j+1}}\right|\\ &= \left|(A^n(x_{j+1})-A^n_{j+1})e^{i\omega\phi^n(x_{j+1})}+A^n_{j+1}e^{i\omega\phi^n(x_{j+1})}(1-e^{i\omega(\phi^n_{j+1}-\phi^n(x_{j+1}))})\right|\\ &\le \left|A^n(x_{j+1})-A^n_{j+1})\right|+|A^n_{j+1}||e^{C\omega\Delta x_c^{p+1}}-1|.\end{aligned}$$

Using Theorem 3, Theorem 5 and Theorem 2, we obtain

$$\begin{aligned}\left|A^n(x_{j+1})-A^n_{j+1})\right| &\le C\Delta x_c^{p+1}\sum_{l=0}^{p+1}|\left(A^n\right)^{(l)}|_{L^\infty([x_j,x_{j+1}])}\le C\Delta x_c^{p+1}|A^n|_{L^\infty([x_j,x_{j+1}])}\\ &= C\Delta x_c^{p+1}|\tilde{v}^n|_{L^\infty([x_j,x_{j+1}])}\le C\Delta x_c^{p+1}|v^n(x_j)|\le C\Delta x_c^{p+1}\delta_0^n,\end{aligned}$$

since from (16) it follows that $|\left(A^n\right)^{(l)}|\le C|A^n|$ for some constant C. Similarly,

$$\begin{aligned}|A^n_{j+1}||e^{C\omega\Delta x_c^{p+1}}-1| &\le C\left(|A^n(x_{j+1})|+|A^n(x_{j+1})-A^n_{j+1}|\right)\omega\Delta x_c^{p+1}\\ &\le C\left(|\tilde{v}^n(x_{j+1})|+\Delta x_c^{p+1}\delta_0^n\right)\omega\Delta x_c^{p+1}\\ &\le C(1+\Delta x_c^{p+1})\omega\Delta x_c^{p+1}\delta_0^n\le C\omega\Delta x_c^{p+1}\delta_0^n.\end{aligned}$$

Hence, together with (62), (63) and (64) we get (61).

Lemma 4 *Assume $\omega>1$ and let J^n be the total number of grid points in G^n. Then*

$$\left|\prod_{\ell=k}^{j}\mathscr{S}^n(x_\ell,\omega)\right|\le\prod_{\ell=0}^{J^n}|\mathscr{S}^n(x_\ell,\omega)|\le C_6,\qquad 0\le k\le j\le J^n.$$

where $\mathscr{S}^n(x_j,\omega)$ is defined by (50) and C_6 is some constant that does not depend on ω.

Proof Let us first show that

$$\left|\mathscr{S}^n(x_j,\omega)\right|\le 1+C\Delta x_f,\qquad x_j\in G_f^n.\tag{65}$$

Apply the R-K scheme to the equation

$$y'(x)=i\omega a(x_j+x,\omega)y(x),\qquad y(0)=1,$$

where $a(x,\omega)$ is defined by (45). Then

$$y_1=1+i\omega\Delta x_f\tilde{a}(x_j,\Delta x_f,\omega)$$

and the exact solution is

$$y(1)=e^{i\omega\int_{x_j}^{x_{j+1}}a(s,\omega)ds}=y_1+\tau_0.$$

Hence,

$$|\mathscr{S}^n(x_j,\omega)| = |1+i\omega\Delta x_f\tilde{a}(x_j,\Delta x_f,\omega)| \leq \left|e^{i\omega\int_{x_j}^{x_{j+1}}a(s,\omega)ds}\right| + |\tau_0| \leq \left|e^{\int_{x_j}^{x_{j+1}}\frac{c_x(s)}{2c(s)}ds}\right| + |\tau_0|$$
$$= \left|e^{\int_{x_j}^{x_{j+1}}d\left(\ln\left(\sqrt{c(s)}\right)\right)}\right| + |\tau_0| \leq \sqrt{\frac{c(x_{j+1})}{c(x_j)}} + |\tau_0| \leq 1 + C\Delta x_f + |\tau_0|, \qquad (66)$$

where C is independent of j. Since we assumed $\omega > 1$ and (54), it follows from Theorem 2 that

$$|\partial_x^p y|_\infty \leq C_p\omega^p,$$

and we can apply Theorem 5 in (66) to obtain that

$$|\tau_0| \leq C(\Delta x_f\omega)^{p+1} \overset{(54)}{<} C\Delta x_f,$$

showing that (65) is true. Note that C_p and C in Theorem 5 are independent of j. For $x_j \in G_c^n$ it follows directly that $\mathscr{S}^n(x_j,\omega) \leq 1 + C\Delta x_c$ with C independent of ω. For $x_j \in G_f^n$, we calculate

$$\prod_{x_j\in G_f^n}\mathscr{S}^n(x_j,\omega) \leq \prod_{x_j\in G_f^n}\left(1+C\Delta x_f\right) \leq \left(1+C\Delta x_f\right)^{J_f} \leq e^{CJ_f\Delta x_f} = e^{CL}.$$

The same calculation can be done for $x_j \in G_c^n$ to obtain

$$\prod_{x_j\in G_c^n}\mathscr{S}^n(x_j,\omega) \leq e^{CJ_c\Delta x_c} = e^{CL}.$$

Taking $C_6 = \exp(2CL)$ the lemma is proved.

Lemma 5 *For $x_j \in G_f^n$,*

$$\left|\tilde{b}^n(x_j,\Delta x,\omega) - \tilde{\bar{b}}^n(x_j,\Delta x,\omega)\right| \leq C_7\begin{cases}\frac{\varepsilon^{n-1}}{\omega} + \max\left((\omega\Delta x_f)^p, \omega\Delta x_c^p)\delta_0^n\right), & n \geq 1\\ 0, & n = 0\end{cases}$$

where C_7 is some constant that does not depend on ω.

Proof The bound is trivial for $n = 0$ by (48). For $n > 0$, it follows from Theorem 5 that

$$\tilde{b}^n(x_j,\omega,\Delta x) = \sum_{k=1}^{s}\tilde{r}_k(x_j,\omega,\Delta x)b^n(x_j+\gamma_k\Delta x),$$

where $b^n(x,\Delta x,\omega)$ is defined by (45). Hence, using (58),

$$\left|\tilde{b}^n(x_j,\Delta x,\omega) - \tilde{\bar{b}}^n(x_j,\Delta x,\omega)\right| \leq \sum_{k=1}^{s}\tilde{r}_k(x_j,\omega,\Delta x)\left|b^n(x_j+\gamma_k\Delta x,\omega) - \bar{b}^n(x_j+\gamma_k\Delta x,\omega)\right|$$
$$\leq \frac{C}{\omega}\sum_{k=1}^{s}\tilde{r}_k(x_j,\omega,\Delta x)\left|f^n(x_j+\gamma_k\Delta x) - \bar{f}^n(x_j+\gamma_k\Delta x)\right|$$
$$\leq \frac{C}{\omega}\left(\varepsilon^{n-1} + \max((\omega\Delta x_f)^p, \omega\Delta x_c^p)\delta_0^{n-1}\right),$$

since $|\tilde{r}_k(x_j,\omega,\Delta x)| \leq C$ by Theorem 5. This proves the lemma.

Next, we prove the following

Theorem 6 *Let ε_v^n be defined by (56). Suppose (51) and (52) hold. Then there exists a $\delta < 1$ such that*

$$\varepsilon^n \le C\delta^n \left(\Delta x_f^p \omega^{p+1} + \Delta x_c^p \omega + \mathrm{Tol} \right). \tag{67}$$

Proof Define $\mathscr{B}^n = \max_j |\tilde{b}^n(x_j) - \tilde{\tilde{b}}^n(x_j)|$. Then, we obtain from Lemma 4

$$\begin{aligned}
\varepsilon_{v(j+1)}^n &\le \left| \mathscr{S}^n(x_j,\omega) v^n(x_j) + \lambda_j^n \Delta x_f \tilde{b}^n(x_j) - \mathscr{S}^n(x_j,\omega) v_j^n - \lambda_j^n \Delta x_f \tilde{\tilde{b}}^n(x_j) \right| + |\tau_j^n| \\
&\le \mathscr{S}^n(x_j,\omega) \varepsilon_{v(j)}^n + \lambda_j^n \Delta x_f \mathscr{B}^n + |\tau_j^n|.
\end{aligned}$$

Let $\mathscr{S}_{k,j}^n(\omega) = \prod_{l=k}^j \mathscr{S}^n(x_l,\omega)$ with $\mathscr{S}_{k+1,k}^n(\omega) := 1$. By induction on the result above,

$$\begin{aligned}
\varepsilon_{v(j+1)}^n &\le \mathscr{S}_{0,j}^n(\omega) \varepsilon_{v(0)}^n + \sum_{k=0}^j \mathscr{S}_{k+1,j}^n(\omega) \left(\lambda_k \Delta x_f \mathscr{B}^n + |\tau_k^n| \right) \\
&\le C_6 \sum_{x_k \in G_f^n} \left(\Delta x_f \mathscr{B}^n + \tau_f^n \right) + C_6 \sum_{x_k \in G_c^n} \tau_c^n \\
&\le C_6 J_f \left(\Delta x_f \mathscr{B}^n + \tau_f^n \right) + C_6 J_c \tau_c^n = \frac{C_6}{\Delta x_f} \left(\Delta x_f \mathscr{B}^n + \tau_f^n \right) + \frac{C_6}{\Delta x_c} \tau_c^n,
\end{aligned}$$

where C_6 is the constant in Lemma 4. This is independent of j and thus true also for ε_v^n. Moreover, the argument can be used for ε_w and therefore the estimate actually holds for ε^n. From Lemma 5 and Lemma 3 for $n = 0$ we obtain

$$\varepsilon^0 \le \tau_f^0 \frac{C_6}{\Delta x_f} + \tau_c^0 \frac{C_6}{\Delta x_c} \le C \left(\Delta x_f^p \omega^{p+1} + \Delta x_c^p \omega + \mathrm{Tol} \right) =: C\left(M + \mathrm{Tol} \right), \tag{68}$$

where we also use the relation $\mathrm{Tol}_n = \omega \mathrm{Tol}/2^{n+1}L$. For $n > 0$, Lemma 5 and Lemma 3 give

$$\begin{aligned}
\varepsilon^n &\le \frac{C_6 C_7}{\omega} \varepsilon^{n-1} + C \max((\omega \Delta x_f)^p, \omega \Delta x_c^p) \delta_0^n + C \Delta x_f^p \omega^{p+1} \delta_0^n \\
&\quad + C_5 \left[\max\left((\Delta x_f \omega)^p, \omega \Delta x_c^p \right) \delta_0^n + \frac{\varepsilon^{n-1} + \mathrm{Tol}_n}{\omega} \right] \\
&\le \frac{C_5 + C_6 C_7}{\omega} \varepsilon^{n-1} + C(M\delta_0^n + \mathrm{Tol} 2^{-n}) \le \tilde{\delta} \varepsilon^{n-1} + C'(M + \mathrm{Tol}) \tilde{\delta}^n,
\end{aligned}$$

where

$$\tilde{\delta} = \max\left(\frac{C_5 + C_6 C_7}{\omega}, \delta_0, \frac{1}{2} \right) < 1,$$

by (51) and (52). We note that there exists a δ such that $\tilde{\delta} < \delta < 1$. Then by induction on the estimate above and using (51), (68), we conclude that (67) holds, since

$$\varepsilon^n \le \tilde{\delta}^n \varepsilon^0 + (M + \mathrm{Tol}) \sum_{k=0}^{n-1} \tilde{\delta}^k \tilde{\delta}^{n-k} = \tilde{\delta}^n \varepsilon^0 + (M + \mathrm{Tol}) n \tilde{\delta}^n \le C\delta^n (M + \mathrm{Tol}).$$

4.4 Proof of Theorem 4

Now, we can finally prove Theorem 4. By Theorem 6 and Theorem 1, there exists a $\delta < 1$ such that

$$\begin{aligned}|u(x) - \tilde{z}^m(x)|_\infty &\le |u(x) - z^m(x)|_\infty + |z^m(x) - \tilde{z}^m(x)|_\infty \\ &\le C(|A| + |f|_{L_1})\delta^{m+1} + \sum_{n=0}^{m} \left(|v^n(x) - \tilde{v}^n(x)|_\infty + |w^n(x) - \tilde{w}^n(x)|_\infty\right). \quad (69)\end{aligned}$$

Using (57)

$$\begin{aligned}\sum_{n=0}^{m} |v^n(x) - \tilde{v}^n(x)|_\infty &\le C \sum_{n=0}^{m} \left(\varepsilon_v^n + \max((\omega\Delta x_f)^p, \omega\Delta x_c^p)\delta_0^n\right) \\ &\le C\left(\left(\Delta x_f^p \omega^{p+1} + \Delta x_c^p \omega + \mathrm{Tol}\right) + \max((\omega\Delta x_f)^p, \omega\Delta x_c^p)\right) \sum_{n=0}^{m} \delta^n \\ &\le C\left(\Delta x_f^p \omega^{p+1} + \Delta x_c^p \omega + \mathrm{Tol}\right).\end{aligned}$$

We get the same estimate for $\sum_{n=0}^{m} |w^n(x) - \tilde{w}^n(x)|_\infty$ and together with (69) we obtain (4).

5 Numerical Experiments

In this section we apply our method to compute the solution of the Helmholtz equation in $[-1,1]$ with different functions $c(x)$. We compare the L_2 error of our method against the L_2 error of a direct finite difference method. We use the finite difference solution with 100 points per wavelength as the exact solution. We also compare the computational cost of the two methods. We compare the total number of grid points in which the solution is computed in the methods and not the direct execution time, in order to factor out the effect of differences in code optimization. In both methods, the cost is directly proportional to this number. Note that the cost is not a smooth function of ω and the general increase in cost with higher ω is far from monotone. In our method, we solve the equations with a 4th order Runge–Kutta method and for the finite difference solution we use a 4th order finite difference scheme. For the interpolation we use cubic (fourth order) spline interpolation. In every experiment, we use

$$\Delta x_f = C_1 \frac{\mathrm{Tol}^{1/4} \min_x c(x)}{\omega^{1+1/4}}, \qquad \Delta x_c = C_2 \omega \mathrm{Tol} \Delta x_f,$$

where the choice of constants C_1 and C_2 depends on the speed function. The same Δx_f is used in the finite difference method.

5.1 Example 1

Let us first consider a nonsymmetric function $c(x)$ that is shown in Figure 3 (left). The absolute value of the solution for $\omega = 16$ is plotted in Figure 3 (right). The oscillations are due to the reflected waves which would not be present in GO. The error, computational cost and the number of iterations needed to compute the solution in our method for $\mathrm{Tol} = 0.1$, is shown in Figure 4. The computational cost here grows essentially as $\omega^{1/4}$ as predicted

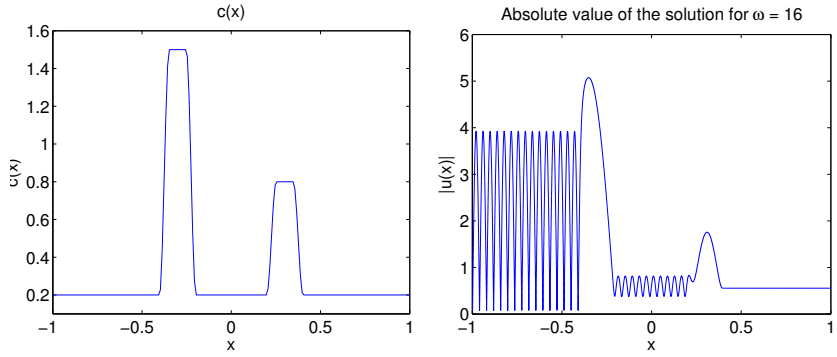

Fig. 3 Function $c(x)$ (left) used in Example 1. Absolute value of the solution for $\omega = 16$ (right).

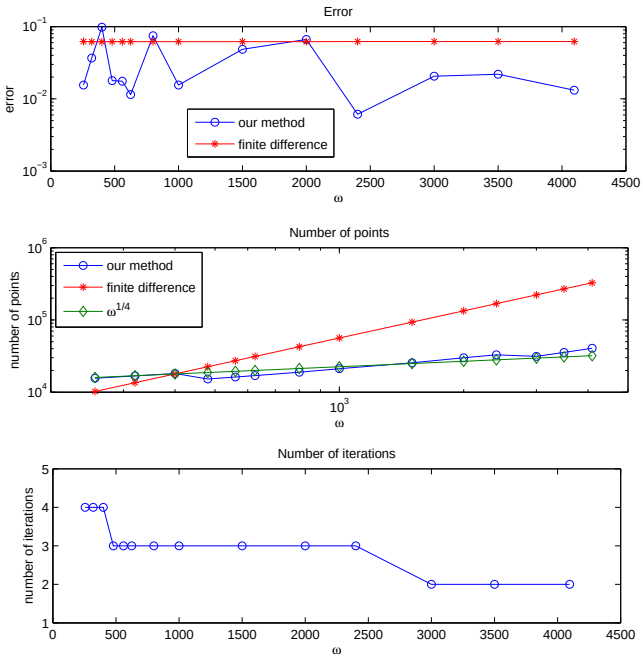

Fig. 4 Comparison between the finite difference method and our method with tolerance Tol $= 0.1$, for $256 \leq \omega \leq 4096$ when the speed function is given in Figure 3. The error (top), the computational cost (middle) and the number of iterations needed to compute the solution with our method (bottom) .

by the asymptotic theory. To illustrate how the forcing function changes in iterations, we choose $\omega = 256$ and plot the part of the function that is larger than Tol_n after the first three iterations in Figure 5. After the fourth iteration, the forcing function is everywhere equal to zero.

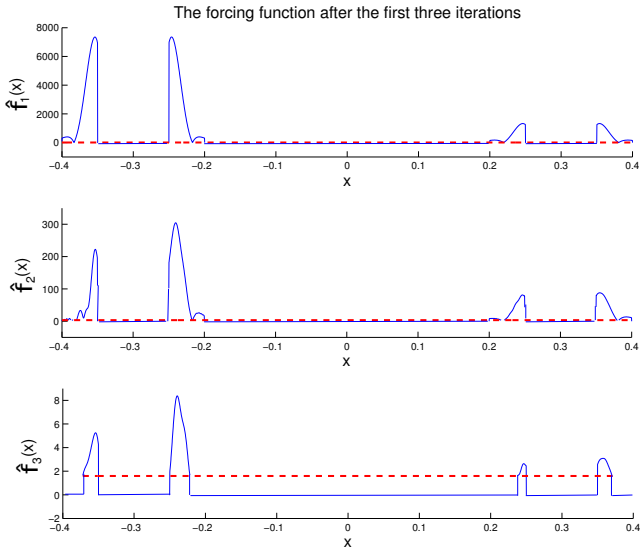

Fig. 5 The forcing function in our method after the first three iterations for $\omega = 256$ when the speed function is given in Figure 3. Everywhere else in the domain, it is equal to zero and we do not plot that part. The dashed line represents Tol_n after every iteration.

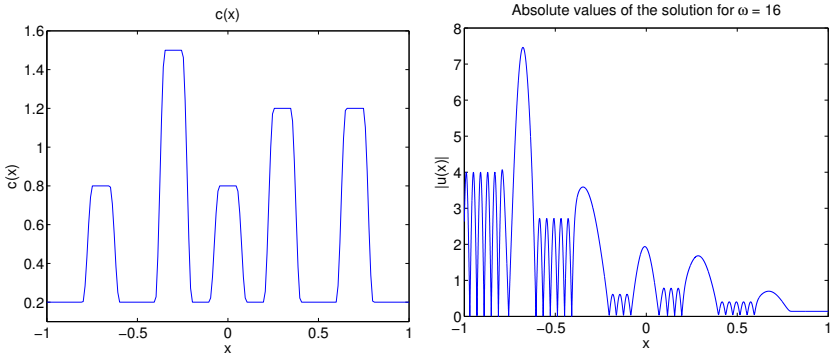

Fig. 6 Function $c(x)$ (left) used in Example 2. Absolute value of the solution for $\omega = 16$ (right).

5.2 Example 2

Let us now consider a more complicated speed function shown in Figure 6 (left). The absolute value of the solution is plotted in Figure 6 (right). The error, computational cost and the number of iterations needed to compute the solution in our method for $\text{Tol} = 0.1$ is shown in Figure 7. For this example the computational cost grows somewhat faster than $\omega^{1/4}$ in the presented range of ω.

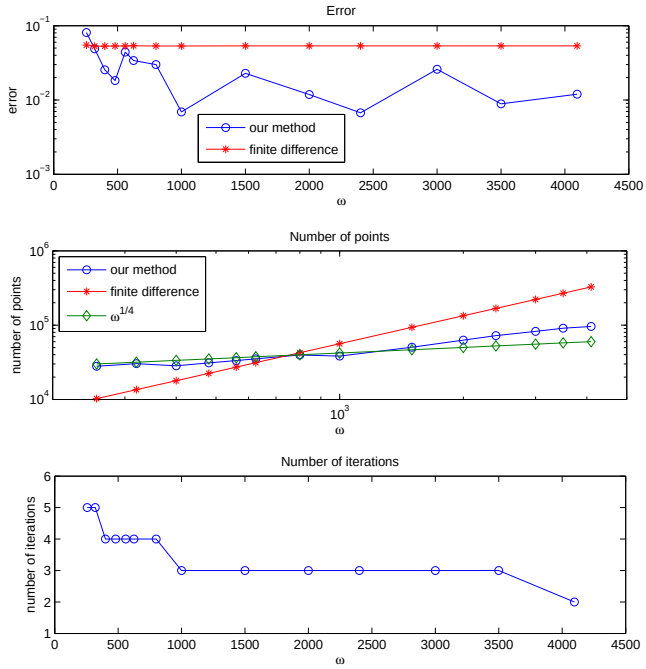

Fig. 7 Comparison between the finite difference method and our method with tolerance Tol $= 0.1$, for $256 \leq \omega \leq 4096$ when the speed function is given in Figure 6. The error (top), the computational cost (middle) and the number of iterations needed to compute the solution with our method (bottom) .

5.3 Example 3

We consider a speed function that is shown in Figure 8 (left). The absolute value of the solution is plotted in Figure 8 (right). The error, computational cost and the number of iterations needed to compute the solution in our method for Tol $= 0.1$ is shown in Figure 9. There is some ambiguity in the rate of incrase for the computational cost, but it appears slightly higher than $\omega^{1/4}$. For this speed function we test the change of number of iterations needed to compute the solution as the frequency decreases. From the theoretical results it follows that ω has to be greater than some number in order to get the prescribed error, c.f. Theorem 1. This means that the number of iterations will increase as ω decreases. The number of iterations plotted as a function of frequency is shown in Figure 10. It can be noted that for ω smaller than a cut-off frequency $\omega_c \approx 50$ the number of iterations needed to compute the solution for Tol $= 0.1$ blows up. This cut-off frequency is much smaller than the theoretical upper bound estimate $\omega_c \leq \beta |\alpha|_{L_1} \approx 550$ given in Theorem 1 for this case. In our experience, this is typical, the method works well also for frequencies significantly lower than the theoretical cut-off frequency.

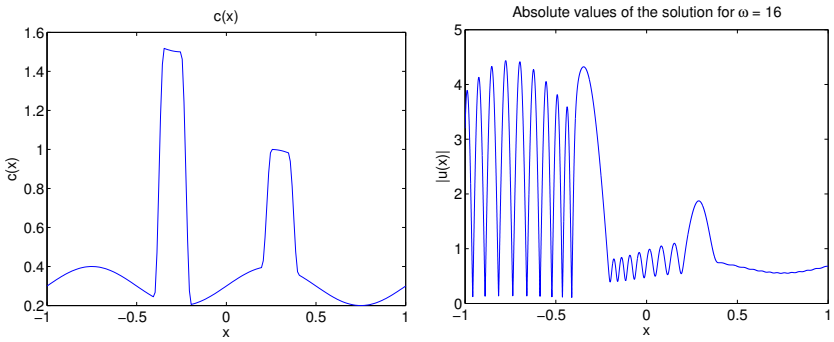

Fig. 8 Function $c(x)$ (left) used in Example 3. Absolute value of the solution for $\omega = 16$ (right).

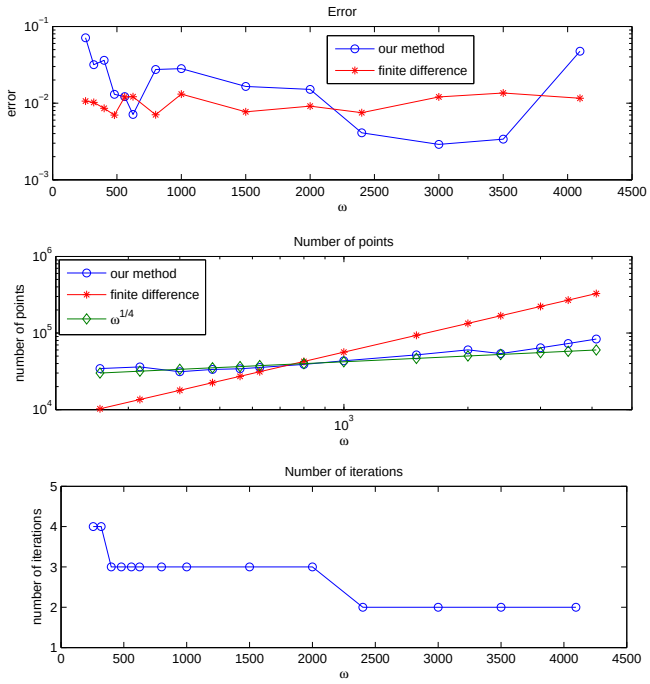

Fig. 9 Comparison between the finite difference method and our method with tolerance Tol $= 0.1$, for $256 \leq \omega \leq 4096$ when the speed function is given in Figure 8. The error (top), the computational cost (middle) and the number of iterations needed to compute the solution with our method (bottom) .

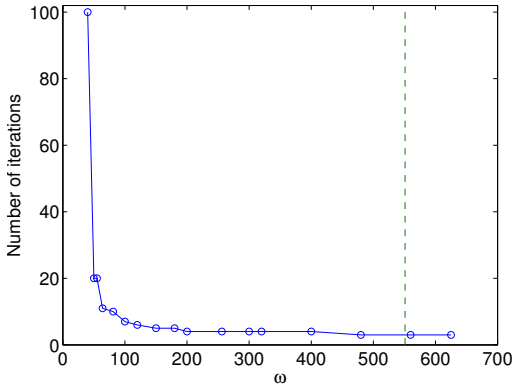

Fig. 10 Number of iterations as a function of ω for $40 \leq \omega \leq 625$, Tol $= 0.1$ and the speed function given in Figure 8. The number of iterations increases rapidly for $\omega < 50$. The dashed line represent the theoretical estimate of the cut-off frequency ω_c.

References

1. Abboud, T., Nédélec, J.C., Zhou, B.: M*é*thode des *é*quations int*é*grales pour les hautes fr*é*quencies. C.R. Acad. Sci. Paris. **318**, 165–170 (1994)
2. Atkinson, F.V.: Wave propagation and the Bremmer series. J. Math. Anal. Appl. **1**, 225–276 (1960)
3. Ayyadi, A.E., Jüngel, A.: Semiconductor simulations using a coupled quantum drift-diffusion Schrdinger-Poisson model. SIAM J. Appl. Math. **66**(2), 554–572 (2006)
4. Babuska, I., Melenk, J.: The partition of unity method. Int. J. Numer. Meths. Eng. **40**, 727–758 (1997)
5. Bellman, R., Kalaba, R.: Functional equations, wave propagation and invariant imbedding. J. Math Mech. **8**, 683–704 (1959)
6. Benamou, J.D.: An introduction to Eulerian geometrical optics (1992-2002). J. Sci. Comput. **19**(1–3), 63–93 (2003)
7. Bremmer, H.: The W.K.B. approximation as the first term of a geometric-optical series. Commun. Pure Appl. Math. **4**, 105–115 (1951)
8. Bruno, O.P., Geuzaine, C.A., Jr., J.A.M., Reitich, F.: Prescribed error tolerances within fixed computational times for scattering problems of arbitrarily high frequency: the convex case. Philos. Transact. A. Math. Phys. Eng. Sci. **362**(1816), 629–645 (2004)
9. Bruno, O.P., Geuzaine, C.A., Reitich, F.: On the $\mathscr{O}(1)$ solution of multiple-scattering problems. IEEE Trans. Magn. **41**, 1488–1491 (2005)
10. Cessenat, O., Despres, B.: Application of an ultra weak variational formulation of elliptic PDEs to the two-dimensional Helmholtz problem. SIAM J. Numer. Anal. **35**(1), 255–299 (1998)
11. Chandler-Wilde, S.N., Graham, I.G.: Boundary integral methods in high frequency scattering. In: B. Engquist, T. Fokas, E. Hairer, A. Iserles (eds.) Highly Oscillatory Problems: Theory, Computation and Applications. Cambridge University Press (2008)
12. Chandler-Wilde, S.N., Langdon, S.: A Galerkin boundary element method for high frequency scattering by convex polygons. SIAM J. Numer. Anal. **45**(2), 610–640 (2007)
13. Červený, V., Molotkov, I.A., Psencik, I.: Ray methods in seismology. Univ. Karlova Press (1977)
14. Dominguez, V., Graham, I.G., Smyshlyaev, V.P.: A hybrid numerical-asymptotic boundary integral method for high-frequency acoustic scattering. Numer. Math. **106**, 471–510 (2007)
15. Ecevit, F., Reitich, F.: Analysis of multiple scattering iterations for high-frequency scattering problems. Numer. Math. **114**(2), 271–354 (2009)
16. Elman, H.C., O'Leary, D.P.: Efficient iterative solution of the three-dimensional Helmholtz equation. J. Comput. Phys. **142**(1), 163–181 (1998)
17. Engquist, B., Runborg, O.: Computational high frequency wave propagation. Acta Numerica **12**, 181–266 (2003)
18. Engquist, B., Runborg, O., Tornberg, A.K.: High-frequency wave propagation by the segment projection method. J. Comput. Phys. **178**, 373–390 (2002)

19. Engquist, B., Ying, L.: Sweeping preconditioner for the Helmholtz equation: Hierarchical matrix representation. Preprint, University of Texas at Austin (2010)
20. Engquist, B., Ying, L.: Sweeping preconditioner for the Helmholtz equation: Moving perfectly matched layers. Preprint, University of Texas at Austin (2010)
21. Erlangga, Y.A.: Advances in iterative methods and preconditioners for the Helmholtz equation. Arch. Comput. Methods Eng. **15**, 37–66 (2008)
22. Erlangga, Y.A., Oosterlee, C., Vuik, C.: A novel multigrid based preconditioner for heterogeneous Helmholtz problem. SIAM J. Sci. Comput. **27**, 1471–1492 (2006)
23. Ernst, O., Golub, G.H.: A domain decomposition approach to solving the Helmholtz equation with a radiating boundary condition. American Mathematical Society, Providence **15**, 177–192 (1994)
24. Fomel, S., Sethian, J.A.: Fast phase space computation of multiple arrivals. Proc. Natl. Acad. Sci. USA **99**(11), 7329–7334 (2002)
25. Giladi, E., Keller, J.B.: A hybrid numerical asymptotic method for scattering problems. Comput. Phys. **174**, 226–247 (2001)
26. Gittelson, C.J., Hiptmair, R., Perugia, I.: Plane wave discontinuous Galerkin methods:Analysis of the h-version. M2AN Math. Model Numer. Anal. **43**, 297–331 (2009)
27. Han, H., Huang, Z.: A tailored finite point method for the Helmholtz equation with high wave numbers in heterogeneous medium. J. Comp. Math. **26**, 728–739 (2008)
28. Huybrechs, D., Vandewalle, S.: A sparse discretisation for integral equation formulations of high-frequency scattering problems. Comm. SIAM J. Sci. Comput. **29**, 2305–2328 (2007)
29. Keller, H.B., Keller, J.B.: Exponential-like solutions of systems of linear ordinary differential equations. J. SIAM **10**, 246–259 (1962)
30. Keller, J.: Geometrical theory of diffraction. J. Opt. Soc. Amer **52** (1962)
31. Langdon, S., Chandler-Wilde, S.N.: A wave number independent boundary element method for an acoustic scattering problem. SIAM J. Num. Anal. **43**, 2450–2477 (2006)
32. Liu, H., Osher, S., Tsai, R.: Multi-valued solution and level set methods in computational high frequency wave propagation. Commun. Comput. Phys. **1**(5), 765–804 (2006)
33. Lundstedt, J., He, S.: Time domain direct and inverse problems for a nonuniform LCRG line with internal sources. IEEE Transactions on Electromagnetic Compatibility **39**(2), 79–88 (1997)
34. Made, M.M.M.: Incomplete factorization-based preconditionings for solving the Helmholtz equation. Int. J. Numer. Methods Eng. **50**, 1077–1101 (2001)
35. Malinen, M., Monk, P., Huttunen, T.: Solving Maxwell's equations using the ultra weak variational formulation. J. Comp. Phys. **223**, 731–758 (2007)
36. McMaken, H.: On the convergence of the Bremmer series for the Helmholtz equation in 2-D. Wave Motion **8**, 277–283 (1986)
37. Monk, P., Huttunen, T.: The use of plane waves to approximate wave propagation in anisotropic media. J. Comp. Math. **25**, 350–36 (2007)
38. Otto, K., Larsson, E.: Iterative solution of the Helmholtz equation by a second–order method. SIAM J. Matrix Anal. Appl. **21**(1), 209–229 (1999)
39. Perrey-Debain, E., Laghrouche, O., Bettess, P., Trevelyan, J.: Plane-wave basis finite elements and boundary elements for three-dimensional wave scattering. Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. **362**(1816), 561–577 (2004)
40. Popov, M.M.: A new method of computation of wave fields using Gaussian beams. Wave Motion **4**, 85–97 (1982)
41. Popovic, J.: A fast method for solving the Helmholtz equation based on wave splitting. Licentiate Thesis, CSC, KTH, Stockhom (2009)
42. Ralston, J.: Gaussian beams and the propagation of singularities. In studies of partial differential equations, Math. Assoc. America, Washington, DC **23**, 206–248 (1982)
43. Runborg, O.: Some new results in multiphase geometrical optics. M2AN Math. Model. Numer. Anal. **34**, 1203–1231 (2000)
44. Runborg, O.: Mathematical models and numerical methods for high frequency waves. Communications in Computational Physics **2**(5), 827–880 (2007)
45. Vinje, V., Iversen, E., Gjøystdal, H.: Traveltime and amplitude estimation using wavefront construction. Geophysics **58**(8), 1157–1166 (1993)
46. Whitham, G.B.: Linear and nonlinear waves. John Wiley & Sons, Inc. (1974)
47. Ying, L., Candès, E.J.: The phase flow method. J. Comput. Phys. **220**(1), 184–215 (2006)