GAUSSIAN BEAM APPROACH FOR THE BOUNDARY VALUE PROBLEM OF HIGH FREQUENCY HELMHOLTZ EQUATION *

CHUNXIONG ZHENG†

Abstract. We propose an asymptotic numerical method called the Gaussian beam approach for the boundary value problem of high frequency Helmholtz equation. The basic idea is to approximate the traveling waves with a summation of Gaussian beams by the least squares algorithm. Gaussian beams are asymptotic solutions of linear wave equations in the high frequency regime. We deduce the ODE systems satisfied by the Gaussian beams up to third order. The key ingredient of the proposed method is the construction of a finite-dimensional beam space which has a good approximating property. If the exact solutions of boundary value problems contain some strongly evanescent wave modes, the Gaussian beam approach might fail. To remedy this problem, we resort to the domain decomposition technique to separate the domain of definition into a boundary layer region and its complementary interior region. The former is handled by a domain-based discretization method, and the latter by the Gaussian beam approach. Schwarz iterations should then be performed based on suitable transmission boundary conditions at the interface of two regions. Numerical tests demonstrate that the proposed method is very promising.

Key words. Gaussian beam, high frequency, Helmholtz equation, domain decomposition, least squares algorithm.

AMS subject classifications. 65N35, 35J05.

1. Introduction

We are aiming at an efficient asymptotic solver for the Helmholtz equation

$$\Delta u + \frac{\omega^2}{c^2(\mathbf{r})} u = 0, \quad \mathbf{r} = (x, y) \in \Omega, \tag{1.1}$$

with some well-posed local linear boundary condition

$$\mathcal{B}u(\mathbf{r}) = f(\mathbf{r}), \quad \mathbf{r} \in \partial\Omega, \tag{1.2}$$

where Ω is a 2-D bounded convex domain on the scale of $O(1)$ and $f(\mathbf{r})$ is a prescribed boundary value function. The velocity field $c(\mathbf{r})$ is assumed to be sufficiently smooth. In general, the solution of the Helmholtz equation (1.1) becomes oscillatory with characteristic wave length $1/\omega$ when the frequency ω is large. To apply a domain-based discretization method, the number of mesh points in each direction should be more than $O(\omega)$ due to the pollution effect of high frequencies [1]. This strong constraint of mesh size makes direct solution of (1.1) very expensive. Numerical methods based on high frequency asymptotics are thus greatly motivated. The readers are referred to [9] for a review.

The classical geometric optics (GO) approach presents a systematic way of finding asymptotic solutions to high frequency linear wave equations. Based on the WKBJ ansatz, the wave function is computed through a phase eikonal equation and a sequence of amplitude transport equations. The eikonal equation can be solved efficiently by the ray tracing method, and the amplitudes are then derived by carrying out some integrals along each specific ray; see [15] for more details.

Despite the great success in handling various wave propagation problems, the classical GO approach has its shortcomings. A significant drawback is that the solution based on the ray tracing does not exist globally. The asymptotic solution ceases to be valid at caustics where the rays intersect and the amplitudes blow up. Many remedies have been proposed to overcome this difficulty, among which is the Gaussian beam (summation) approach; see [5, 3, 21] for examples. In a way similar to the GO approach, the Gaussian beam approach also resorts to the concept of ray tracing: the central curve of each Gaussian beam is exactly a specific ray. The difference lies in the value space of phase function. While always real in the classical GO method, the phase function of Gaussian beam is generally complex in its valid region. Since the imaginary part of the phase Hessian projected to the transverse direction is positive definite and bounded from below, the Gaussian beam is always well-defined and bell-shaped. This property ensures the validity of the Gaussian beam approach even at caustics. In [20] the approximating accuracy of the Gaussian beam solution has been studied theoretically.

In this paper we propose an asymptotic solver based on the Gaussian beam theory for the boundary value problem (1.1)–(1.2) in the high frequency regime. The basic idea is to approximate the traveling waves with a summation of Gaussian beams. We determine the approximate solution by solving a least squares problem in a finite-dimensional beam space, which is spanned by a set of Gaussian beams. In two dimensions, our solver involves only $O(\omega)$ unknowns, much less than $O(\omega^2)$ for any domain-based discretization method. When the exact solution contains some strongly evanescent wave modes, we employ the domain decomposition technique to divide the domain of definition into a boundary layer evanescent wave region and its complementary traveling wave region. The evanescent wave region is discretized by finite elements, and the traveling wave region is handled by the proposed Gaussian beam approach. Schwarz iterations are performed based on suitable transmission boundary conditions at the interface of two regions. Since the width of evanescent wave region is at most $O(\omega^{-1/2})$, the number of overall unknowns is $O(\omega^{3/2})$, which is still a gain compared with any direct spatial discretization method.

The organization of the rest is as follows. In section 2, we apply the Gaussian beam theory to the specific Helmholtz equation (1.1), and explicitly deduce the ODE systems satisfied by the beam solutions up to third order. In section 3, we make an error analysis on the Gaussian beam approximations for the plane waves. Section 4 presents the Gaussian beam approach for the boundary value problems. In section 5, we demonstrate the effectiveness of the proposed method with some test problems whose solutions mainly consist of traveling waves. The Gaussian beam approach combined with the domain decomposition technique is considered in section 6 for general boundary value problems whose solutions might contain some strongly evanescent wave modes. Section 7 concludes this paper and briefly discusses the advantages and disadvantages of the proposed method.

2. Gaussian beams

In the high frequency regime, it is a common practice to seek asymptotic solutions to (1.1) by using the WKBJ ansatz

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

Here the phase function ϕ and the amplitude functions A_k are assumed to be smooth and independent of the frequency ω. Substituting the above ansatz into the Helmholtz

equation (1.1) and equating the coefficients of different powers with respect to ω, we derive a series of PDEs

$$\nabla\phi\cdot\nabla\phi = 1/c^2(\mathbf{r}), \tag{2.1}$$

$$2\nabla\phi\cdot\nabla A_0 + \Delta\phi A_0 = 0, \tag{2.2}$$

$$2\nabla\phi\cdot\nabla A_{k+1} + \Delta\phi A_{k+1} + \Delta A_k = 0, \ k = 0, 1, \ldots. \tag{2.3}$$

In the classical GO approach, the phase function ϕ is real. Given suitable boundary condition, the phase function ϕ can be determined by the following ray-tracing system

$$\frac{dx}{dt} = c(\mathbf{r})\cos\theta, \ \frac{dy}{dt} = c(\mathbf{r})\sin\theta, \ \frac{d\theta}{dt} = c_x(\mathbf{r})\sin\theta - c_y(\mathbf{r})\cos\theta, \tag{2.4}$$

where t is the optical length parameter and θ is the angle of ray direction. After that, the amplitudes A_k can be computed through the transport equation (2.2)–(2.3). The readers are referred to [15] for details.

The main drawback of the classical GO approach is that the asymptotic solution ceases to be valid at caustics where the rays intersect and the amplitudes blow up. Considering this point, many authors proposed the Gaussian beam approach; see [5, 3, 21] for examples. The Gaussian beam approach also resorts to the ray tracing system: the beam center is exactly a single ray. The difference between the classical GO approach and the Gaussian beam approach is that while the phase function is real-valued in the former, it is generally complex-valued for each specific Gaussian beam in the latter. Projected to the transverse directions, the imaginary part of the phase Hessian is positive definite, which ensures that the transverse profile of wave function is always bell-shaped.

The Gaussian beam construction dates back to many years ago and has been considered by many authors. Here we do not intend for even any rough review. The readers are referred to [4] for more information. As to the specific Helmholtz equation (1.1), a detailed construction and mathematical analysis can be found in principle in the Appendix part of [23]. However, for the sake of completeness of this paper, we reconsider this issue in the rest of this section.

For the N-th order ($N \geq 1$) Gaussian beam, equations (2.1)–(2.3) do not hold rigorously but only asymptotically, i.e.,

$$\nabla\phi\cdot\nabla\phi - 1/c^2(\mathbf{r}) = O(\omega^{-1-N/2}), \tag{2.5}$$

$$2\nabla\phi\cdot\nabla A_0 + \Delta\phi A_0 = O(\omega^{-N/2}), \tag{2.6}$$

$$2\nabla\phi\cdot\nabla A_{k+1} + \Delta\phi A_{k+1} + \Delta A_k = O(\omega^{k+1-N/2}), \ k = 0, 1, \ldots. \tag{2.7}$$

The governing Helmholtz equation thus holds to $O(\omega^{1-N/2})$, and the beam solution has an accuracy of $O(\omega^{-N/2})$. Note that for $k \geq N/2 - 1$, equation (2.7) holds trivially.

In the ray-centered right-hand coordinates system (n, t), where n is the transverse distance from the beam center, we have

$$\nabla = \frac{\mathbf{e}_t}{h}\partial_t + \mathbf{e}_n\partial_n, \quad \Delta = \frac{1}{h}\left(\frac{1}{h}\right)_t \partial_t + \frac{1}{h^2}\partial_t^2 + \frac{c_n}{h}\partial_n + \partial_n^2.$$

Here $\mathbf{e}_t = (\cos\theta, \sin\theta)^\top$, $\mathbf{e}_n = (\sin\theta, -\cos\theta)^\top$, $h = c + nc_n$, $c \equiv c(\mathbf{r}(t))$, $c_n \equiv \mathbf{e}_n\cdot\nabla c(\mathbf{r}(t))$ and $\mathbf{r}(t)$ denotes the central ray. The left hand sides of (2.5)–(2.7) are reformulated

as

$$\nabla\phi\cdot\nabla\phi-1/c^2(\mathbf{r})=\frac{1}{h^2}\phi_t^2+\phi_n^2-1/c^2(\mathbf{r}), \tag{2.8}$$

$$2\nabla\phi\cdot\nabla A_0+\Delta\phi A_0=\frac{2}{h^2}\phi_t A_{0,t}+2\phi_n A_{0,n}$$

$$+\frac{1}{h}\left(\frac{1}{h}\right)_t\phi_t A_0+\frac{1}{h^2}\phi_{tt}A_0+\frac{c_n}{h}\phi_n A_0+\phi_{nn}A_0, \tag{2.9}$$

$$2\nabla\phi\cdot\nabla A_{k+1}+\Delta\phi A_{k+1}+\Delta A_k=\frac{2}{h^2}\phi_t A_{k+1,t}+2\phi_n A_{k+1,n}$$

$$+\frac{1}{h}\left(\frac{1}{h}\right)_t\phi_t A_{k+1}+\frac{1}{h^2}\phi_{tt}A_{k+1}+\frac{c_n}{h}\phi_n A_{k+1}+\phi_{nn}A_{k+1}$$

$$+\frac{1}{h}\left(\frac{1}{h}\right)_t A_{k,t}+\frac{1}{h^2}A_{k,tt}+\frac{c_n}{h}A_{k,n}+A_{k,nn}. \tag{2.10}$$

Performing Taylor expansion with respect to n for those variable coefficient functions in (2.8)–(2.10) yields

$$\frac{1}{h^2}=b_0+b_1 n+b_2 n^2+b_3 n^3+b_4 n^4+\dots, \tag{2.11}$$

$$\frac{1}{c^2(\mathbf{r})}=p_0+p_1 n+p_2 n^2+p_3 n^3+p_4 n^4+\dots, \tag{2.12}$$

$$\frac{1}{h}\left(\frac{1}{h}\right)_t=q_0+q_1 n+q_2 n^2+\dots, \tag{2.13}$$

$$\frac{c_n}{h}=r_0+r_1 n+\dots, \tag{2.14}$$

where

$$b_0=\frac{1}{c^2}, \; b_1=-\frac{2c_n}{c^3}, \; b_2=\frac{3c_n^2}{c^4}, \; b_3=-\frac{4c_n^3}{c^5}, \; b_4=\frac{5c_n^4}{c^6},$$

and

$$p_0=\frac{1}{c^2}, \; p_1=-\frac{2c_n}{c^3}, \; p_2=\frac{3c_n^2}{c^4}-\frac{c_{nn}}{c^3}, \; p_3=-\frac{4c_n^3}{c^5}+\frac{3c_n c_{nn}}{c^4}-\frac{c_{nnn}}{3c^3},$$

$$p_4=\frac{5c_n^4}{c^6}-\frac{6c_n^2 c_{nn}}{c^5}+\frac{3c_{nn}^2}{4c^4}+\frac{c_n c_{nnn}}{c^4}-\frac{c_{nnnn}}{12c^3},$$

and

$$q_0=-\frac{c_s}{c^2}, \; q_1=-\frac{c_{ns}}{c^2}+\frac{3c_n c_s}{c^3}, \; q_2=\frac{3c_n c_{ns}}{c^3}-\frac{6c_n^2 c_s}{c^4},$$

and

$$r_0=\frac{c_n}{c}, \; r_1=-\frac{c_n^2}{c^2}.$$

Here c_{nn}, c_{nnn}, c_{nnnn}, c_s, and c_{ns} denote different order (mixed) derivatives of the velocity field valued at the central ray curve $\mathbf{r}(t)$.

Suppose the Taylor expansion of the phase function with respect to n is given by

$$\phi=t+\frac{\phi_2}{2}n^2+\frac{\phi_3}{6}n^3+\frac{\phi_4}{24}n^4+\dots, \tag{2.15}$$

where ϕ_2, ϕ_3 and ϕ_4 are functions only of t. Substituting (2.15), (2.11)–(2.12) into (2.8) and equating the different powers of n (notice that $n = O(\omega^{-1/2})$ for the Gaussian beams), we realize that the eikonal equation (2.5) holds trivially within $O(1)$ and $O(n)$. For higher orders we have

$$O(n^2): \qquad b_0\phi_{2,t} + \phi_2^2 + b_2 - p_2 = 0, \tag{2.16}$$

$$O(n^3): \qquad \frac{b_0\phi_{3,t}}{3} + \phi_{2,t}b_1 + b_3 + \phi_2\phi_3 - p_3 = 0, \tag{2.17}$$

$$O(n^4): \qquad \frac{b_0\phi_{4,t}}{12} + \frac{b_0\phi_{2,t}^2}{4} + \frac{\phi_{3,t}}{3}b_1 + \phi_{2,t}b_2 + b_4 + \frac{\phi_3^2}{4} + \frac{\phi_2\phi_4}{3} - p_4 = 0. \tag{2.18}$$

Equation (2.16) is a simple Ricatti equation. It may be reduced by standard techniques to a pair of first-order linear equations. But from the computational point of view, this is unnecessary. Note that if ϕ_2 has a positive imaginary part at the starting point, it also does at any advancing point.

Next we come to the amplitude functions A_k. Suppose

$$A_0 = A_{00} + A_{01}n + \frac{A_{02}}{2}n^2 + \dots. \tag{2.19}$$

Substituting (2.19), (2.11), (2.13)-(2.14) into (2.9) and equating the different powers of n gives

$$O(n^0): \quad 2b_0 A_{00,t} + q_0 A_{00} + \phi_2 A_{00} = 0, \tag{2.20}$$

$$O(n^1): \quad 2b_0 A_{01,t} + 2b_1 A_{00,t} + 3\phi_2 A_{01} + q_0 A_{01} + q_1 A_{00} + r_0\phi_2 A_{00} + \phi_3 A_{00} = 0, \tag{2.21}$$

$$O(n^2): \quad b_0 A_{02,t} + b_0\phi_{2,t}A_{00,t} + 2b_2 A_{00,t} + 2b_1 A_{01,t} + 2\phi_3 A_{01} + 5\phi_2 A_{02}/2$$
$$+ q_0\phi_{2,t}A_{00}/2 + q_2 A_{00}/2 + q_1 A_{01} + q_0 A_{02}/2 + b_0\phi_{2,tt}A_{00}/2$$
$$+ r_0\phi_3 A_{00}/2 + \phi_2(r_0 A_{01} + r_1 A_{00}) + \phi_4 A_{00}/2 = 0. \tag{2.22}$$

Analogously, for the equation (2.7) with $k=0$, within $O(1)$ the amplitude function A_1 satisfies

$$2b_0 A_{1,t} + q_0 A_1 + \phi_2 A_1 + q_0 A_{00,t} + b_0 A_{00,tt} + r_0 A_{01} + A_{02} = 0. \tag{2.23}$$

Higher order Taylor expansion terms can be also considered in principle, but the formulations would be much more complicated.

The first three Gaussian beams are

$$u_{GB1} = A_{00}\exp\left\{i\omega\left(t + \frac{\phi_2}{2}n^2\right)\right\}, \tag{2.24}$$

$$u_{GB2} = (A_{00} + A_{01}n)\exp\left\{i\omega\left(t + \frac{\phi_2}{2}n^2 + \frac{\phi_3}{6}n^3\right)\right\}, \tag{2.25}$$

$$u_{GB3} = \left(A_{00} + A_{01}n + \frac{A_{02}}{2}n^2 + \frac{1}{i\omega}A_1\right)\exp\left\{i\omega\left(t + \frac{\phi_2}{2}n^2 + \frac{\phi_3}{6}n^3 + \frac{\phi_4}{24}n^4\right)\right\}. \tag{2.26}$$

Given the initial data

$$\phi_2(0) = a_2, \ \phi_3(0) = a_3, \ \phi_4(0) = a_4, A_{00}(0) = a_{00}, \ A_{01}(0) = a_{01}, \ A_{02}(0) = a_{02}, \ A_1(0) = a_1, \tag{2.27}$$

the first order Gaussian beam can be computed by solving the ray-tracing system (2.4) together with the equations (2.16) and (2.20), and the second order with equations

(2.17), (2.21), and the third order with equations (2.18), (2.22), and (2.23). In the case of constant velocity $c(\mathbf{r}) \equiv 1$, the ray is simply a straight line, and other ODEs can be integrated out analytically. See the results given in Appendix A. In the general case, if the ray curve is expressed in an analytical or semi-analytical way, all phase functions and amplitude functions can be integrated out analytically, as performed in [17, 18]. In this paper we simply employ the fourth-order Runge-Kutta scheme to simultaneously compute the ray curve and the related phase and amplitude functions.

Any Gaussian beam has its *valid* region. This is typically because the ray-centered coordinates system can only be set up locally unless the central ray is a straight line. The valid region of Gaussian beam is determined by the curvature of its central ray. Thus to justify the Gaussian beam approximation, the beam solution should be sufficiently small at the boundary of valid region. This is always possible if the frequency ω is large enough. In this paper, we set the valid region of Gaussian beam as $\{(n,t) \in \mathbb{R}^2 | \omega n^2 \operatorname{Im} \phi_2(t)/2 < 36\}$.

3. Error analysis for the beam approximation to plane waves

In this section, under the assumption of constant velocity field $c(\mathbf{r}) \equiv 1$, we give analysis for a special type of beam summation approximation to the plane wave solution of the form

$$u(\lambda, \mathbf{r}) = \exp\left(i\lambda\sqrt{\omega}\, x\right) \cdot \exp\left(i\sqrt{\omega^2 - \lambda^2 \omega}\, y\right), \tag{3.1}$$

where $\lambda \in \mathbb{C}$ is a free parameter on the scale of $O(1)$, and the square root $\sqrt{\cdot}$ is taken with a positive real part. Two interesting cases are included:

I if λ is real, the solution (3.1) represents a propagating wave. The propagating direction is almost parallel to the positive y-direction for large ω. When λ runs through a symmetric interval with a length of $O(1)$, the wave propagating angles will cover an interval centered at $\frac{\pi}{2}$ with a length of $O(\omega^{-1/2})$. For example, if λ is valued in $[-1,1]$, the covered range of angle is

$$[\pi/2 - \arctan(1/\sqrt{\omega - 1}), \pi/2 + \arctan(1/\sqrt{\omega - 1})].$$

Asymptotically, the length of this interval is $2\omega^{-1/2}$.

II if λ is not real, the solution (3.1) is evanescent in a direction almost parallel to the x-axis. The evanescent direction angle covers an interval with a length of $O(\omega^{-1/2})$, if the real part of λ runs through an interval of length $O(1)$.

Denote $\phi_\omega^N(\mathbf{r})$ to be the N-th order Gaussian beam propagating in the positive y-direction with the initial data $\phi_\omega(x) = \exp(-\omega x^2/2)$ at $y = 0$. Set $x_j = jh$ with h being the spatial stepsize, and the Gaussian beam approximation u_{GB}^N as

$$u_{GB}^N(\lambda, \mathbf{r}) = \sum_{k \in \mathbb{Z}} a_k(\lambda)\phi_{\omega,k}^N(\mathbf{r}), \tag{3.2}$$

where

$$a_k(\lambda) = \frac{\exp(i\lambda\sqrt{\omega}\, x_k)}{\sum_{k \in \mathbb{Z}} \exp(i\lambda\sqrt{\omega}\, x_k)\phi_\omega(x_k)}, \quad \phi_{\omega,k}^N(\mathbf{r}) = \phi_\omega^N(x - x_k, y).$$

We would like to study under what conditions the relative error $(u_{GB}^N - u)/u$ can be made small. Note that u_{GB}^N satisfies the interpolating condition

$$u_{GB}^N(\lambda, x_k, 0) = \exp\left(i\lambda\sqrt{\omega}\, x_k\right) = u(\lambda, x_k, 0), \quad \forall k \in \mathbb{Z}. \tag{3.3}$$

THEOREM 3.1. *Given $y_{max} > 0$, $\lambda_{rm} > 0$, $\lambda_{im} \geq 0$, set $\lambda_m = \sqrt{\lambda_{rm}^2 + \lambda_{im}^2}$, and let a_0 be the constant satisfying (3.17). Then for any $\omega \geq \max\left(1, \lambda_m^4 y_{max}/2\right)$, any $h \leq \omega^{-1/2}/a^*$ with*

$$a^* = \max\left(a_0, \frac{2(\lambda_{rm} + y_{max}\lambda_{im})}{\pi}\right),$$

any λ satisfying $|\mathrm{Re}\,\lambda| \leq \lambda_{rm}$, and $|\mathrm{Im}\,\lambda| \leq \lambda_{im}$, the following holds:

$$\left|\frac{u_{GB}^1(\lambda, \mathbf{r}) - u(\lambda, \mathbf{r})}{u(\lambda, \mathbf{r})}\right| \leq 0.33 \exp\left(-\frac{\pi^2}{\omega h^2}\right) + \lambda_m^4 y_{max} \exp\left(\lambda_{rm}^2/2\right)/\omega,$$

$$\forall \mathbf{r} = (x, y) \text{ with } |y| \leq y_{max}. \tag{3.4}$$

The following proof is essentially based on the analysis in [16].

Proof.   Since

$$\frac{u_{GB}^1(\lambda, \mathbf{r})}{u(\lambda, \mathbf{r})} = \frac{\sum_{k \in \mathbb{Z}} \exp(i\lambda\sqrt{\omega}(x_k - x))\phi_\omega^1(x - x_k, y)\exp\left(-i\sqrt{\omega^2 - \lambda^2 w}\,y\right)}{\sum_{k \in \mathbb{Z}} \exp(i\lambda\sqrt{\omega}x_k)\phi_\omega(x_k)},$$

by setting

$$\psi(\lambda, \mathbf{r}) = \sum_{k \in \mathbb{Z}} \exp(i\lambda\sqrt{\omega}(x_k - x))\phi_\omega^1(x - x_k, y)\exp\left(-i\sqrt{\omega^2 - \lambda^2 w}\,y\right), \tag{3.5}$$

we have

$$\psi(\lambda, x, 0) = \sum_{k \in \mathbb{Z}} \exp(i\lambda\sqrt{\omega}(x_k - x))\phi_\omega(x - x_k),$$

and

$$u_{GB}^1(\lambda, \mathbf{r})/u(\lambda, \mathbf{r}) = \psi(\lambda, \mathbf{r})/\psi(\lambda, 0, 0). \tag{3.6}$$

It is straightforward to verify (using the results in Appendix A) that the first order beam solution ϕ_ω^1 satisfies the Schrödinger equation

$$\phi_{\omega, y}^1 = i\omega\left(1 + \frac{\partial_x^2}{2\omega^2}\right)\phi_\omega^1.$$

Thus ψ (see (3.5)) satisfies

$$\psi_y = i\left(\omega - \sqrt{\omega^2 - \lambda^2\omega} + \frac{\partial_x^2 + 2i\lambda\sqrt{\omega}\partial_x - \lambda^2\omega}{2\omega}\right)\psi.$$

Since ψ is h-periodic with respect to the variable x, we expand it into a Fourier series

$$\psi(\lambda, \mathbf{r}) = \sum_{k \in \mathbb{Z}} b_k(\lambda, y)\exp(2\pi ikx/h), \tag{3.7}$$

where the k-th Fourier coefficient b_k satisfies the relation

$$b_k(\lambda, y) = b_k(\lambda, 0)\exp\left(iy\left[\omega - \sqrt{\omega^2 - \lambda^2\omega} - \frac{4\pi^2k^2/h^2 + 4\pi\lambda\sqrt{\omega}k/h + \lambda^2\omega}{2\omega}\right]\right)$$

$$= b_k(\lambda, 0)\exp\left(iy\left[\omega - \sqrt{\omega^2 - \lambda^2\omega} - \frac{\lambda^2}{2} - 2\pi^2k^2a^2 - 2\pi\lambda ka\right]\right). \tag{3.8}$$

Here we have set $a = \omega^{-1/2}/h$. The initial value

$$b_k(\lambda,0) = \frac{1}{h}\int_0^h \psi(\lambda,x,0)\exp(-2\pi i k x/h)\,dx$$

may be evaluated analytically as

$$
\begin{aligned}
b_k(\lambda,0) &= \frac{1}{h}\sum_{j\in\mathbb{Z}}\int_0^h \exp(i\lambda\sqrt{\omega}(x_j-x))\phi_\omega(x-x_j)\exp(-2\pi i k x/h)\,dx\\
&= \frac{1}{h}\sum_{j\in\mathbb{Z}}\int_{x_{-j}}^{x_{1-j}} \exp(-i\lambda\sqrt{\omega}x)\phi_\omega(x)\exp(-2\pi i k(x+x_j)/h)\,dx\\
&= \frac{1}{h}\sum_{j\in\mathbb{Z}}\int_{x_{-j}}^{x_{1-j}} \exp(-i\lambda\sqrt{\omega}x)\phi_\omega(x)\exp(-2\pi i k x/h)\,dx\\
&= \frac{1}{h}\int_{-\infty}^{\infty} \exp(-i\lambda\sqrt{\omega}x)\phi_\omega(x)\exp(-2\pi i k x/h)\,dx\\
&= \frac{\sqrt{2\pi}}{h\sqrt{\omega}}\exp\left(-\frac{(\lambda\sqrt{\omega}+\frac{2\pi k}{h})^2}{2\omega}\right) = \sqrt{2\pi}\,a\exp\left(-\frac{(\lambda+2\pi k a)^2}{2}\right). \quad (3.9)
\end{aligned}
$$

Since

$$|1+\sqrt{z}|\ge 1,\quad \forall z\in\mathbb{C},$$

we have

$$
\left|\omega-\sqrt{\omega^2-\lambda^2\omega}-\frac{\lambda^2}{2}\right| = \left|\frac{\lambda^2\omega}{\omega+\sqrt{\omega^2-\lambda^2\omega}}-\frac{\lambda^2}{2}\right| = \left|\frac{\lambda^2(\omega-\sqrt{\omega^2-\lambda^2\omega})}{2(\omega+\sqrt{\omega^2-\lambda^2\omega})}\right|
$$

$$
= \left|\frac{\lambda^4\omega}{2(\omega+\sqrt{\omega^2-\lambda^2\omega})^2}\right| = \left|\frac{\lambda^4}{2\omega\left(1+\sqrt{1-\lambda^2/\omega}\right)^2}\right| \le \frac{\lambda_m^4}{2\omega}. \quad (3.10)
$$

Thus (see (3.8), (3.9)),

$$
\begin{aligned}
&|b_k(\lambda,y)|\\
&= \left|\sqrt{2\pi}\,a\exp\left(-\frac{(\lambda+2\pi k a)^2}{2}\right)\exp\left(iy\left[\omega-\sqrt{\omega^2-\lambda^2\omega}-\frac{\lambda^2}{2}-2\pi^2 k^2 a^2-2\pi\lambda k a\right]\right)\right|\\
&\le \sqrt{2\pi}\,a\exp\left(\frac{(\mathrm{Im}\,\lambda)^2-(\mathrm{Re}\,\lambda+2\pi k a)^2}{2}\right)\exp\left(\left(\frac{\lambda_m^4}{2\omega}+2\pi k a\,\mathrm{Im}\,\lambda\right)|y|\right). \quad (3.11)
\end{aligned}
$$

According to the result in Appendix C, if $\lambda_m^4 y_{max}/(2\omega)\le 1$, i.e., $\omega\ge\lambda_m^4 y_{max}/2$, we have

$$
\begin{aligned}
|b_0(\lambda,y)-b_0(\lambda,0)| &= |b_0(\lambda,0)|\left|\exp\left(iy\left[\omega-\sqrt{\omega^2-\lambda^2\omega}-\frac{\lambda^2}{2}\right]\right)-1\right|\\
&\le \frac{0.86\lambda_{max}^4 y_{max}}{\omega}\sqrt{2\pi}\,a\exp\left(\frac{(\mathrm{Im}\,\lambda)^2}{2}\right). \quad (3.12)
\end{aligned}
$$

Consulting (3.7) and (3.11) we have

$$
|\psi(\lambda,\mathbf{r}) - b_0(\lambda,y)| = \left| \sum_{k\in\mathbb{Z},k\neq 0} b_k(\lambda,y)\exp(2\pi ikx/h) \right| \leq \sum_{k\in\mathbb{Z},k\neq 0} |b_k(\lambda,y)|
$$

$$
\leq \sqrt{2\pi}\,a \sum_{k\in\mathbb{Z},k\neq 0} \exp\left(\frac{(\operatorname{Im}\lambda)^2 - (\operatorname{Re}\lambda + 2\pi ka)^2}{2} \right) \exp\left(\left(\frac{\lambda_m^4}{2\omega} + 2\pi ka\operatorname{Im}\lambda \right)|y| \right)
$$

$$
= \sqrt{2\pi}\,a\exp\left(\frac{(\operatorname{Im}\lambda)^2}{2} + \frac{\lambda_m^4|y|}{2\omega} \right) \sum_{k\in\mathbb{Z},k\neq 0} \exp\left(-\frac{(\operatorname{Re}\lambda + 2\pi ka)^2}{2} \right) \exp(2\pi ka|y|\operatorname{Im}\lambda)
$$

$$
= \sqrt{2\pi}\,a\exp\left(\frac{(\operatorname{Im}\lambda)^2}{2} + \frac{\lambda_m^4|y|}{2\omega} - \frac{|y|\operatorname{Im}\lambda(2\operatorname{Re}\lambda - |y|\operatorname{Im}\lambda)}{2} \right)
$$

$$
\sum_{k\in\mathbb{Z},k\neq 0} \exp\left(-\frac{(\operatorname{Re}\lambda - |y|\operatorname{Im}\lambda + 2\pi ka)^2}{2} \right).
$$

If $\omega \geq 1$ and $(\lambda_{rm} + y_{max}\lambda_{im})/(2\pi a) \leq 1/4$, applying the result in Appendix B yields

$$
|\psi(\lambda,\mathbf{r}) - b_0(\lambda,y)|
$$
$$
\leq \sqrt{2\pi}\,a\exp\left(\frac{(\operatorname{Im}\lambda)^2}{2} + \frac{\lambda_m^4|y|}{2\omega} - \frac{|y|\operatorname{Im}\lambda(2\operatorname{Re}\lambda - |y|\operatorname{Im}\lambda)}{2} \right) \frac{2\exp(-9\pi^2 a^2/8))}{1 - \exp(-5\pi^2 a^2)}
$$
$$
\leq \sqrt{2\pi}\,a\exp\left(\frac{(\operatorname{Im}\lambda)^2}{2} + \frac{\lambda_m^4 y_{max}}{2} + \frac{y_{max}\lambda_{im}(2\lambda_{rm} + y_{max}\lambda_{im})}{2} \right) \frac{2\exp(-9\pi^2 a^2/8))}{1 - \exp(-5\pi^2 a^2)}
$$
$$
= \sqrt{2\pi}\,a\exp\left(\frac{(\operatorname{Im}\lambda)^2}{2} \right) \epsilon(a,\lambda_{rm},\lambda_{im},y_{max}), \tag{3.13}
$$

where we have set

$$
\epsilon(a,\lambda_{rm},\lambda_{im},y_{max})
$$
$$
= \exp\left(\frac{\lambda_m^4 y_{max}}{2} + \frac{y_{max}\lambda_{im}(2\lambda_{rm} + y_{max}\lambda_{im})}{2} \right) \frac{2\exp(-9\pi^2 a^2/8))}{1 - \exp(-5\pi^2 a^2)}. \tag{3.14}
$$

In terms of (3.9) and (3.13), we have

$$
|\psi(\lambda,0,0)| \geq |b_0(\lambda,0)| - |\psi(\lambda,0,0) - b_0(\lambda,0)|
$$
$$
\geq \sqrt{2\pi}\,a\exp\left(\frac{(\operatorname{Im}\lambda)^2 - (\operatorname{Re}\lambda)^2}{2} \right) - \sqrt{2\pi}\,a\exp\left(\frac{(\operatorname{Im}\lambda)^2}{2} \right) \epsilon(a,\lambda_{rm},\lambda_{im},y_{max})
$$
$$
\geq \sqrt{2\pi}\,a\exp\left(\frac{(\operatorname{Im}\lambda)^2}{2} \right) \left(\exp\left(-\frac{\lambda_{rm}^2}{2} \right) - \epsilon(a,\lambda_{rm},\lambda_{im},y_{max}) \right). \tag{3.15}
$$

Moreover (see (3.13) and (3.12)),

$$
|\psi(\lambda,\mathbf{r}) - \psi(\lambda,0,0)|
$$
$$
\leq |\psi(\lambda,\mathbf{r}) - b_0(\lambda,y)| + |\psi(\lambda,0,0) - b_0(\lambda,0)| + |b_0(\lambda,y) - b_0(\lambda,0)|
$$
$$
\leq 2\sqrt{2\pi}\,a\exp\left(\frac{(\operatorname{Im}\lambda)^2}{2} \right) \epsilon(a,\lambda_{rm},\lambda_{im},y_{max}) + \frac{0.86\lambda_m^4 y_{max}}{\omega}\sqrt{2\pi}\,a\exp\left(\frac{(\operatorname{Im}\lambda)^2}{2} \right).
$$
$$
\tag{3.16}
$$

Recalling (3.6) and applying (3.15) and (3.16) we derive

$$\frac{|u_{GB}^1(\lambda,\mathbf{r})-u(\lambda,\mathbf{r})|}{|u(\lambda,\mathbf{r})|}=\frac{|\psi(\lambda,\mathbf{r})-\psi(\lambda,0,0)|}{|\psi(\lambda,0,0)|}\leq\frac{2\epsilon(a,\lambda_{rm},\lambda_{im},y_{max})+0.86\lambda_m^4 y_{max}/\omega}{\exp\left(-\frac{\lambda_{rm}^2}{2}\right)-\epsilon(a,\lambda_{rm},\lambda_{im},y_{max})}.$$

Let a_0 be the constant satisfying

$$\exp\left(\frac{\lambda_{rm}^2}{2}\right)\exp\left(\frac{\lambda_m^4 y_{max}}{2}+\frac{y_{max}\lambda_{im}(2\lambda_{rm}+y_{max}\lambda_{im})}{2}\right)\frac{2\exp(-\pi^2 a^2/8))}{1-\exp(-5\pi^2 a^2)}=0.14, \tag{3.17}$$

then for any $a\geq a_0$, i.e., $h\leq\omega^{-1/2}/a_0$, we have

$$\frac{|u_{GB}^1(\lambda,\mathbf{r})-u(\lambda,\mathbf{r})|}{|u(\lambda,\mathbf{r})|}\leq\frac{2\exp\left(\frac{\lambda_{rm}^2}{2}\right)\epsilon(a,\lambda_{rm},\lambda_{im},y_{max})+0.86\exp\left(\frac{\lambda_{rm}^2}{2}\right)\lambda_m^4 y_{max}/\omega}{1-\exp\left(\frac{\lambda_{rm}^2}{2}\right)\epsilon(a,\lambda_{rm},\lambda_{im},y_{max})}$$

$$\leq\frac{0.28\exp(-\pi^2 a^2)+0.86\exp\left(\frac{\lambda_{rm}^2}{2}\right)\lambda_m^4 y_{max}/\omega}{0.86}$$

$$\leq 0.33\exp\left(-\frac{\pi^2}{\omega h^2}\right)+\exp\left(\frac{\lambda_{rm}^2}{2}\right)\lambda_m^4 y_{max}/\omega.$$

This ends the proof.   □

Two remarks would be made about Theorem 3.1.

REMARK 3.2. The parameter $a=\omega^{-1/2}/h$ reflects the spatial resolution of Gaussian beam summation. As revealed in Theorem 3.1, the spatial stepsize h should sufficiently resolve the characteristic width $\omega^{-1/2}$ of each beam. Note that the first term on the right hand side of (3.4) decays extremely fast with respect to a. For example, by setting

$$\lambda_{rm}=1,\ \lambda_{im}=1,\ y_{max}=1,\ a=1.4,$$

we then have

$$0.33\exp\left(-\frac{\pi^2}{\omega h^2}\right)<1.32\times 10^{-9}.$$

This implies that in a large enough frequency range, the relative error is first order with respect to the frequency ω for the first Gaussian beam approximation.

REMARK 3.3.   The upper bound of the error estimate in Theorem 3.1 is optimal with respect to ω, but the constant multiplied with ω^{-1} can be made smaller. The optimal value should be close to

$$\lambda_m^4 y_{max}\exp\left(\lambda_{rm}^2/2\right)/8\omega,$$

for sufficiently large ω. See (3.10) and (3.12).

It is tempting to extend the result in Theorem 3.1 to higher order beam approximations. The second order is the same as the first order in the constant velocity case, thus there is no need to consider it. For the third order approximation, there exists one essential difficulty which prevents us from making an analogous analysis. Since

$\phi_4(y) = -3y/(1+iy)^4$ (see Appendix A) and the quartic term dominates in the phase function for large x, the beam solution is not valid in the whole x-axis. It should be either confined into a finite domain around the beam center, as we do in this paper, or modified by multiplying a smooth cut-off function which remains 1 near the center and decays to 0 sufficiently fast away from the center. This makes the Fourier analysis (like (3.9)) much more complicated.

At the end of this section, we make some numerical tests on the approximating error of Gaussian beam approximation (3.2). We set $a = 1.4$, and the test region of λ as a 2-by-2 square centered at the origin. Figure 3.1 shows the maximum relative errors of Gaussian beam approximation in the domain $\mathbb{R} \times [-1,1]$ for four typical choices of λ. One could see that the error for the third order approximation is second order with respect to ω, and the error for the first order is first order. An interesting point is that at $\lambda = 0$, the error for the first order approximation saturates for any frequency in the considered range. This is also consistent with our error analysis given in Theorem 3.1: since the second term with respect to ω in (3.4) in fact vanishes, the error is roughly the interpolating error (see (3.3)), which does not change with respect to ω if we fix $a = \omega^{-1/2}/h$.
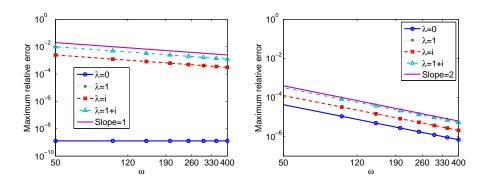

Fig. 3.1. *Maximum relative error.* $h = \omega^{-1/2}/a$ *with* $a = 1.4$. *Left: first order. Right: third order.*

Theorem 3.1 together with the above numerical tests shows that when $a = \omega^{-1/2}/h$ is sufficiently large, the error mainly depends on the frequency ω. For the sake of simplicity in the following, we always assume

HYPOTHESIS 3.4. *The spatial stepsize* h *is small enough and the following error estimate holds*

$$\left| \frac{u_{GB}^N(\lambda, \mathbf{r}) - u(\lambda, \mathbf{r})}{u(\lambda, \mathbf{r})} \right| \leq c\omega^{-\gamma_N}, \quad \forall \mathbf{r} \in \Omega, \ \forall \lambda \in \mathbb{C} \ with \ |\lambda| \leq \lambda_m, \tag{3.18}$$

with $\gamma_N = 1$ *for* $N = 1$ *and* $\gamma_N = 2$ *for* $N = 3$. *Here and hereafter, we use the notation* c *to denote a constant depending on* λ_m *and the size of definition domain* Ω, *but not on* ω. *It may have different values at different places.*

4. Gaussian beam approach for boundary value problems

Now we come to the boundary value problem (1.1)-(1.2) with $c(\mathbf{r}) \equiv 1$. We introduce

$$W_\omega(\Omega) \stackrel{def}{=} \{u_\omega \in H^2(\Omega) : (\Delta + \omega^2)u_\omega = 0\}$$

as the solution space to the Helmholtz equation. For any $g_\omega \in L^1_{per}([0, 2\pi])$ we define the *Herglotz wave function* as

$$u(g_\omega, \mathbf{r}) = \int_0^{2\pi} \exp(i\omega(\cos\theta x + \sin\theta y)) g_\omega(\theta) d\theta. \qquad (4.1)$$

Here g_ω is called the density function. It has been proved (see Theorem 2 in [25]) that within any given error tolerance, any sufficiently smooth solution to the constant frequency Helmholtz equation on a finite domain of smooth boundary can be approximated by a Herglotz function. Considering our problem, if the exact solution u_ω (here we explicitly show the dependence on ω) can be extended within the same scale from Ω to a larger smooth domain $\tilde{\Omega}$, it is reasonable to assume that u_ω can be approximated in Ω by a Herglotz function $u(g_\omega, \mathbf{r})$ uniformly with respect to ω. More precisely, we assume that:

HYPOTHESIS 4.1. *There exist two constants α and β such that for any sufficiently small $\epsilon_0 > 0$ and any large enough ω, there is a function $g_\omega \in L^1_{per}[0, 2\pi]$ satisfying*

$$||u_\omega(\mathbf{r}) - u(g_\omega, \mathbf{r})||_{L^2(\Omega)} \le \epsilon_0, \quad ||g_\omega||_{L^1} = O(\omega^\alpha), \quad ||u_\omega(\mathbf{r})||_{L^2(\Omega)} = O(\omega^\beta). \qquad (4.2)$$

We give several examples here:

1. Plane wave solution $u_\omega(\mathbf{r}) = \exp(i\omega(\cos\theta_0 x + \sin\theta_0 y))$ with a propagating angle θ_0. The solution itself is a Herglotz function with the density function $g_\omega(\theta) = \delta(\theta - \theta_0)$. We have $\alpha = \beta = 0$.

2. Cylindrical wave solution $u_\omega(\mathbf{r}) = J_0(\omega r)$. It is a Herglotz function with the plane wave expansion

$$J_0(\omega r) = \frac{J_0(0)}{2\pi} \int_0^{2\pi} \exp(i\omega(\cos\theta x + \sin\theta y)) d\theta.$$

We have $\alpha = 0$. The asymptotic behavior of $J_0(\omega r)$ for large ω gives $\beta = -1/2$.

3. Point wave solution $u_\omega(\mathbf{r}) = H_0^{(1)}(\omega r)$ where $r = |\mathbf{r}|$ and $H_0^{(1)}$ is the zero-th order Hankel function of the first kind. Without loss of generality, we suppose the convex domain Ω is in the upper half plane. According to [6, Equ. 7], we have

$$H_0^{(1)}(\omega r)$$
$$= \frac{1}{\pi} \left(-2i \int_0^\infty \exp(i\omega x \cosh v - \omega y \sinh v) dv + \int_0^\pi \exp(i\omega x \cos\theta + i\omega y \sin\theta) d\theta \right),$$
$$\forall y > 0.$$

For large ω, the first term in the bracket decays exponentially. Thus we have $\alpha = 0$. The asymptotic behavior of $H^{(1)}(\omega r)$ for large ω gives $\beta = -1/2$.

The Herglotz wave function (4.1) is an integral summation of plane waves. If the support of g_ω is contained in an interval I of length $O(\omega^{-1/2})$ centered at $\pi/2$, we have (see Hypothesis 3.4)

$$\left| u(g_\omega, \mathbf{r}) - \int_I u_{GB}^N(\sqrt{\omega}\cos\theta, \mathbf{r}) g_\omega(\theta) d\theta \right|$$
$$= \left| \int_I \exp(i\omega(\cos\theta x + \sin\theta y)) g_\omega(\theta) d\theta - \int_I u_{GB}^N(\sqrt{\omega}\cos\theta, \mathbf{r}) g_\omega(\theta) d\theta \right|$$
$$= \left| \int_I \left(\exp(i\omega(\cos\theta x + \sin\theta y)) - u_{GB}^N(\sqrt{\omega}\cos\theta, \mathbf{r}) \right) g_\omega(\theta) d\theta \right| \le c\omega^{-\gamma_N} \int_I |g_\omega(\theta)| d\theta, \ \forall \mathbf{r} \in \Omega,$$
$$(4.3)$$

because

$$|\sqrt{\omega}\cos\theta| = |\sqrt{\omega}\sin(\pi/2 - \theta)| = O(1), \ \forall \theta \in [\theta_s, \theta_e].$$

Note that (see (3.2))

$$\int_I u_{GB}^N(\sqrt{\omega}\cos\theta, \mathbf{r}) g_\omega(\theta) d\theta = \sum_{k \in \mathbb{Z}} \phi_{\omega,k}^N(\mathbf{r}) \int_I a_k(\sqrt{\omega}\cos\theta) g_\omega(\theta) d\theta.$$

The error estimate (4.3) thus implies that one can find an element $u(\mathbf{r}) \in$ Span$\{\phi_{\omega,k}^N(\mathbf{r})\}_{k \in \mathbb{Z}}$ which satisfies

$$||u(g_\omega, \mathbf{r}) - u(\mathbf{r})||_{L^2(\Omega)} \leq c\omega^{-\gamma_N} \int_I |g_\omega(\theta)| d\theta. \tag{4.4}$$

In the general case, we could divide the interval $[0, 2\pi]$ (in fact the unit circle) into M sub-intervals I_j with same length of $O(\omega^{-1/2})$. This means $M = O(\omega^{1/2})$. The center of I_j is set as $\pi/2 + 2\pi j/M$, $j = 0, 1, \ldots, M-1$. The Herglotz function $u(g_\omega, \mathbf{r})$ could then be expressed as

$$u(g_\omega, \mathbf{r}) = \sum_{j=0}^{M-1} u(\chi_{I_j} g_\omega, \mathbf{r}),$$

where χ is the indicator function. Since each $u(\chi_{I_j} g_\omega, \mathbf{r})$ is a Herglotz function with a density function supported in I_j, analogous to (4.4), one could find an element $u_j(\mathbf{r})$ in the space Span$\{\phi_{\omega,k}^N(\mathcal{R}_j \mathbf{r})\}_{k \in \mathbb{Z}}$ ($\mathcal{R}_j$ denotes the rotation transformation with an angle of $2\pi j/M$), such that

$$||u(\chi_{I_j} g_\omega, \mathbf{r}) - u_j(\mathbf{r})||_{L^2(\Omega)} \leq c\omega^{-\gamma_N} \int_{I_j} |g_\omega(\theta)| d\theta.$$

Thus (see (4.2)),

$$\left\| u(g_\omega, \mathbf{r}) - \sum_{j=0}^{M-1} u_j(\mathbf{r}) \right\|_{L^2(\Omega)} \leq \sum_{j=0}^{M-1} ||u(\chi_{I_j} g_\omega, \mathbf{r}) - u_j(\mathbf{r})||_{L^2(\Omega)}$$

$$\leq c\omega^{-\gamma_N} \int_0^{2\pi} |g_\omega(\theta)| d\theta \leq c\omega^{\alpha - \gamma_N},$$

and

$$\left\| u_\omega(\mathbf{r}) - \sum_{j=0}^{M-1} u_j(\mathbf{r}) \right\|_{L^2(\Omega)} \leq ||u_\omega(\mathbf{r}) - u(g_\omega, \mathbf{r})||_{L^2(\Omega)} + \left\| u(g_\omega, \mathbf{r}) - \sum_{j=0}^{M-1} u_j(\mathbf{r}) \right\|_{L^2(\Omega)}$$

$$\leq \epsilon_0 + c\omega^{\alpha - \gamma_N}.$$

Setting ϵ_0 small enough, we thus know that there exists an element $\tilde{u}_\omega(\mathbf{r}) = \sum_{j=0}^{M-1} u_j(\mathbf{r})$ in the *beam space*

$$\Phi \stackrel{def}{=} \text{Span}\{\phi_{\omega,k}^N(\mathcal{R}_j \mathbf{r})\}_{j=0,\ldots,M-1, k \in \mathbb{Z}}, \tag{4.5}$$

such that for any large enough ω,

$$\|u_\omega(\mathbf{r}) - \tilde{u}_\omega(\mathbf{r})\|_{L^2(\Omega)} \leq c\omega^{\alpha-\gamma_N}, \tag{4.6}$$

and (see (4.2))

$$\frac{\|u_\omega(\mathbf{r}) - \tilde{u}_\omega(\mathbf{r})\|_{L^2(\Omega)}}{\|u_\omega(\mathbf{r})\|_{L^2(\Omega)}} \leq c\omega^{\alpha-\beta-\gamma_N}. \tag{4.7}$$

THEOREM 4.2. *Under Hypothesis 4.1, there exists a constant $\gamma \leq \alpha - \beta$ such that for any large enough ω, there is an element $\tilde{u}_\omega(\mathbf{r}) \in \Phi$ satisfying*

$$\frac{\|u_\omega(\mathbf{r}) - \tilde{u}_\omega(\mathbf{r})\|_{L^2(\Omega)}}{\|u_\omega(\mathbf{r})\|_{L^2(\Omega)}} \leq c\omega^{\gamma-\gamma_N}. \tag{4.8}$$

REMARK 4.3.   Our numerical tests in the next section suggest that the lower bound of γ may be smaller than $\alpha - \beta$. This implies that the analysis (4.7) on the approximating potential of Φ may not be optimal.

REMARK 4.4.   The beam space Φ (4.5) is infinite-dimensional. Since Ω is a bounded domain on the scale of $O(1)$, and $\phi_{\omega,k}^N(\mathbf{r})$ indeed vanishes on Ω for large $|k|$ (numerically the transverse support of $\phi_{\omega,k}^N(\mathbf{r})$ is on the scale of $O(\omega^{-1/2})$), the function $\tilde{u}_\omega$ satisfying (4.6) is in fact in a finite-dimensional subspace of Φ. Only those beams whose support intersects with Ω have real approximating effect. See figure 4.1. Practically, we determine a priori a circle of radius R which encloses Ω, and draw $M = O(\omega^{1/2})$ straight lines with equidistant direction angles. On each line, we put $[2Ra\omega^{1/2}]$ equidistant points including two ending points. Associated with each point, we compute an N-th order Gaussian beam solution which propagates in a direction perpendicular to the located line. The total number of beam solutions is thus $M \times [2Ra\omega^{1/2}] = O(\omega)$. Notice that figure 4.1 actually shows 10 lines, not 5.

The approximating potential analysis (4.8) ensures us an element $\tilde{u}_\omega(\mathbf{r}) \in \Phi$ which is a good asymptotic approximation to the exact solution if $\gamma < \gamma_N$. However, we could not determine such an element by the above-explained constructive method. Moreover, we aim to develop a general solver for the Helmholtz equation even with a variable velocity field. In this case the analysis based on the plane wave expansion does not work any more. Instead, we determine a function $u_\omega^*(\mathbf{r}) \in \Phi$ which solves the following optimization problem:

$$||\mathcal{B}u_\omega^*(\mathbf{r}) - f(\mathbf{r})||_{L^2(\partial\Omega)} = \min_{v \in \Phi} ||\mathcal{B}v(\mathbf{r}) - f(\mathbf{r})||_{L^2(\partial\Omega)}. \tag{4.9}$$

Since

$$||\mathcal{B}u_\omega^*(\mathbf{r}) - f(\mathbf{r})||_{L^2(\partial\Omega)} \leq ||\mathcal{B}\tilde{u}_\omega(\mathbf{r}) - f(\mathbf{r})||_{L^2(\partial\Omega)},$$

we have

$$||\mathcal{B}u_\omega^*(\mathbf{r}) - \mathcal{B}\tilde{u}_\omega(\mathbf{r})||_{L^2(\partial\Omega)} \leq ||\mathcal{B}u_\omega^*(\mathbf{r}) - f(\mathbf{r})||_{L^2(\partial\Omega)} + ||\mathcal{B}\tilde{u}_\omega(\mathbf{r}) - f(\mathbf{r})||_{L^2(\partial\Omega)}$$

$$\leq 2||\mathcal{B}\tilde{u}_\omega(\mathbf{r}) - f(\mathbf{r})||_{L^2(\partial\Omega)} = 2||\mathcal{B}\tilde{u}_\omega(\mathbf{r}) - \mathcal{B}u_\omega(\mathbf{r})||_{L^2(\partial\Omega)}$$

$$\leq 2||\mathcal{B}|| \cdot ||u_\omega(\mathbf{r}) - \tilde{u}_\omega(\mathbf{r})||_{L^2(\Omega)}.$$
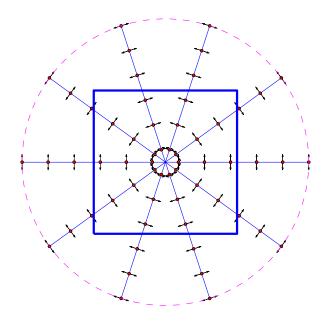
FIG. 4.1. *Schematic view of beam waist centers and propagating directions. The closed blue curve stands for the boundary of convex domain of definition.*

Thus,

$$||u_\omega^*(\mathbf{r}) - \tilde{u}_\omega(\mathbf{r})||_{L^2(\Omega)} \le ||\mathcal{B}^{-1}|| \cdot ||\mathcal{B}u_\omega^*(\mathbf{r}) - \mathcal{B}\tilde{u}_\omega(\mathbf{r})||_{L^2(\partial\Omega)}$$

$$\le 2||\mathcal{B}^{-1}|| \cdot ||\mathcal{B}|| \cdot ||u_\omega(\mathbf{r}) - \tilde{u}_\omega(\mathbf{r})||_{L^2(\Omega)},$$

and

$$||u_\omega(\mathbf{r}) - u_\omega^*(\mathbf{r})||_{L^2(\Omega)} \le ||u_\omega^*(\mathbf{r}) - \tilde{u}_\omega(\mathbf{r})||_{L^2(\Omega)} + ||\tilde{u}_\omega(\mathbf{r}) - u_\omega(\mathbf{r})||_{L^2(\Omega)}$$

$$\le (1 + 2||\mathcal{B}^{-1}|| \cdot ||\mathcal{B}||) ||u_\omega(\mathbf{r}) - \tilde{u}_\omega(\mathbf{r})||_{L^2(\Omega)}.$$

If $||\mathcal{B}^{-1}|| \cdot ||\mathcal{B}|| = O(1)$, we derive (see (4.8))

$$\frac{||u_\omega(\mathbf{r}) - u_\omega^*(\mathbf{r})||_{L^2(\Omega)}}{||u_\omega(\mathbf{r})||_{L^2(\Omega)}} \le c \frac{||u_\omega(\mathbf{r}) - \tilde{u}_\omega(\mathbf{r})||_{L^2(\Omega)}}{||u_\omega(\mathbf{r})||_{L^2(\Omega)}} \le c\omega^{\gamma - \gamma_N}. \qquad (4.10)$$

We conclude the above analysis as a theorem.

THEOREM 4.5. *Assume that Hypothesis 4.1 holds and*

1. *$\mathcal{B}$ is a bounded operator from $W_\omega + \Phi \subset L^2(\Omega)$ to $L^2(\partial\Omega)$;*

2. *Confined to Φ, $\mathcal{B}$ is injective. Moreover, the associated inverse operator $\mathcal{B}^{-1}$ is bounded from $\mathcal{B}\Phi \subset L^2(\partial\Omega)$ to $\Phi \subset L^2(\Omega)$;*

3. *$||\mathcal{B}|| \cdot ||\mathcal{B}^{-1}|| = O(1)$.*

Then the error estimate (4.10) holds.

Several remarks would be made here.

REMARK 4.6.   The first two assumptions made in Theorem 4.5 are not hard to verify if $\mathcal{B}$ is a well-posed local linear boundary operator at most of first order. The point is the third assumption. Take the Dirichlet boundary operator as an example. We could expect $||\mathcal{B}||=O(1)$. $||\mathcal{B}^{-1}||$ typically reflects the stability of the Gaussian beam formulation. For simple geometries, it is possible to derive $||\mathcal{B}^{-1}||=O(1)$ by the Fourier analysis conducted in the proof of Theorem 3.1. For a more general geometry, the analysis is much more complicated.

REMARK 4.7.   The error estimate (4.10) implies that if $\gamma-\gamma_N<0$, the solution $u_\omega^*(\mathbf{r})$ to the least squares problem (4.9) is a valid asymptotic solution. The case of $\gamma>0$ generally implies that the solution $u_\omega(\mathbf{r})$ contains some strongly evanescent wave modes. If $u_\omega(\mathbf{r})$ mainly consists of traveling waves, we could expect $\gamma=0$. In this case, our method might be asymptotically optimal.

Now let us consider the numerical issues related to the least squares problem (4.9). Let $\{\varphi_j(\mathbf{r})\}_{j=1}^L$ be the effective part of Gaussian beams spanning Φ (see Remark 4.4). We express the solution of (4.9) as $u_\omega^*(\mathbf{r})=\sum_{j=1}^L a_j^*\varphi_j(\mathbf{r})$. The problem (4.9) is then changed into:

$$\left\|\sum_{j=1}^L a_j^*\mathcal{B}\varphi_j(\mathbf{r})-f(\mathbf{r})\right\|_{L^2(\partial\Omega)} = \min_{\{a_j\}_{j=1}^L\in\mathbf{C}^L}\left\|\sum_{j=1}^L a_j\mathcal{B}\varphi_j(\mathbf{r})-f(\mathbf{r})\right\|_{L^2(\partial\Omega)}. \qquad (4.11)$$

This problem involves the L^2 norm at the boundary $\partial\Omega$. In the numerical implementation, it has to be approximated by a suitable quadrature rule. Suppose $\{\mathbf{r}_j\}_{j=1}^{N_p}$ ($N_p=O(\omega)$) is a set of points uniformly distributed at the boundary $\partial\Omega$ (if possible). Using the trapezoidal rule yields a discrete counterpart of (4.11):

$$||A\vec{a}^*-b||_2 = \min_{\vec{a}\in\mathbf{C}^L}||A\vec{a}-b||_2, \qquad (4.12)$$

where $A=(a_{ij})$ is an $N_p\times L$ matrix with $a_{ij}=\mathcal{B}\varphi_j(\mathbf{r}_i)$, $b=(b_i)$ is an N_p-dimensional vector with $b_i=f(\mathbf{r}_i)$, and $||\cdot||_2$ is the standard vector l^2 norm. Other quadrature rules could be used alternatively, which result in a weighted discrete l^2 norm, but the solution strategy would be essentially the same.

As explained in Remark 4.4, the column number of A is $O(\omega)$. Since the beam width is $O(\omega^{-1/2})$, a typical column of A has only $O(\omega^{1/2})$ nonzero elements (see the numerical study on the size of N_p in the next section). If ω is large, iterative solvers based on Krylov subspace projection are thus preferable. Since A is generally ill-conditioned, some regularization technique should be applied to regularize the least squares problem (4.12). The regularization technique for ill-posed linear problems has been a hot research area for several decades [10, 12, 14]. However, we did not find a method in the existing literature which solves our problem very efficiently. Fortunately, our numerical tests reveal that the following algorithm based on the projection regularization works fairly well:

• Solve the normal equation

$$A^*A\vec{a}=A^*b, \qquad (4.13)$$

with CGLS [13, 2, 22] to derive $\vec{a}^*=\{a_j^*\}_{j=1}^L$;

- The iteration process is terminated if the relative residual norm is less than a prescribed tolerance ϵ_0, or the maximum iteration number k_{max} is reached;
- The least squares solution is set as $u_\omega^*(\mathbf{r}) = \sum_{j=1}^L a_j^* \varphi_j(\mathbf{r})$.

In the numerical tests considered in this paper, we set $\epsilon_0 = 10^{-12}$ and $k_{max} = 2000$.

5. Numerical tests

In this section, we present some numerical tests to demonstrate the performance of the proposed Gaussian beam approach. The domain of definition Ω is set as a unit square $[-0.5, 0.5]^2$, and the boundary operator is $\mathcal{B} = \partial_\nu - i\omega$, where ν denotes the outer normal direction of Ω. In all numerical tests, the radius R of the circle which encloses Ω is 1 (see 4.1), and the spatial stepsize is set as $h = 2R/\lceil 2R\omega^{1/2}a \rceil$ with $a = 1.4$. Here $\lceil \cdot \rceil$ denotes the smallest integer which is not less than the argument.
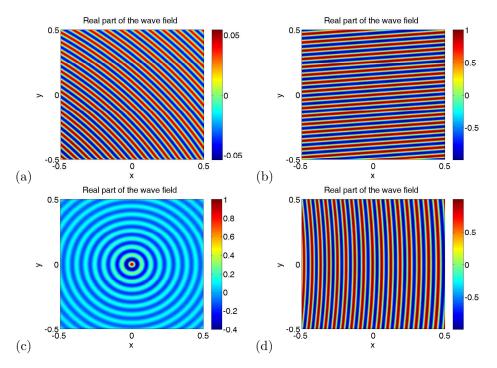

FIG. 5.1. *Real part of the exact wave field.* $\omega = 100$. *(a): point wave. (b): plane wave with* $\theta_0 = \frac{\pi}{2} + \frac{\pi}{4\lceil \sqrt{\omega} \rceil}$. *(c): cylindrical wave. (d): variable velocity solution with* $a = 5$.

We consider four different problems:

Point wave: The exact solution is $u(\mathbf{r}) = H_0^{(1)}\left(\omega\sqrt{(x+2)^2 + (y+2)^2}\right)$, where $H_0^{(1)}$ is the zeroth order Hankel function of the first kind. The point source is located at $(-2, -2)$.

Plane wave: The exact solution is $u(\mathbf{r}) = \exp\left(i\omega(\cos\theta_0 x + \sin\theta_0 y)\right)$. The parameter θ_0 is the propagating directional angle.

Cylindrical wave: The exact solution is $u(\mathbf{r}) = J_0(\omega r)$. From the GO point of view, the origin is a caustics point. The classical GO method could not solve this solution in the whole domain of definition.
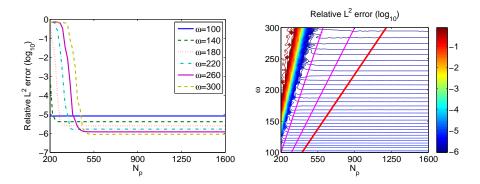
FIG. 5.2. *Dependence of the approximating error on the number of boundary points. Third order Gaussian beam approach is used.* $M = 4\lceil\sqrt{\omega}\rceil$. *Left: Error plot for several* ω. *Right: Contour of errors for* $(N_p,\omega) \in [200, 1600] \times [100, 300]$. *The three lines (from left to right) are* $N_p = 2\omega$, $N_p = 3\omega$ *and* $N_p = 4\omega$, *respectively.*
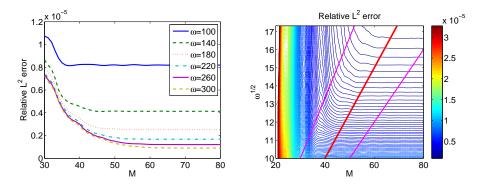

FIG. 5.3. *Dependence of the approximating error on the number of straight lines. Third order Gaussian beam approach is used.* $N_p = 4\lceil\omega\rceil$. *Left: Error plot for several* ω. *Right: Contour of errors for* $(M,\omega) \in [20, 80] \times [100, 300]$. *The three lines (from left to right) are* $M = 3\omega^{1/2}$, $M = 4\omega^{1/2}$ *and* $M = 5\omega^{1/2}$, *respectively.*

Variable velocity problem: The exact solution is

$$u(\mathbf{r}) = \exp\left\{ ia\omega \ln\left(\sqrt{(x+a)^2 + y^2} \right) \right\}, \quad c(\mathbf{r}) = \frac{1}{a}\sqrt{(x+a)^2 + y^2}.$$

The parameter $a > 0.5$ controls the fluctuation of the velocity field.

For ease of reference, we plot the real part of these wave fields in figure 5.1 for $\omega = 100$.

We first make a numerical study on the size of N_p, the number of collocated boundary points. Since the solution is oscillatory of characteristic length ω^{-1}, it is natural to conjecture that the number of boundary points N_p should at least be proportional to ω. In figure 5.2 we show the numerical errors of the point wave solution on the line $x = 0$ for different ω and different N_p by fixing the number of straight lines as $M = 4\lceil\omega^{1/2}\rceil$. From the left error plot in figure 5.2, one can see that the error decreases significantly at the initial stage and then remains stable for large enough N_p. A good estimate of the breakpoints is $N_p = 2\omega$, see the right contour plot in figure 5.2. This means that on average 2π points are necessary and sufficient to
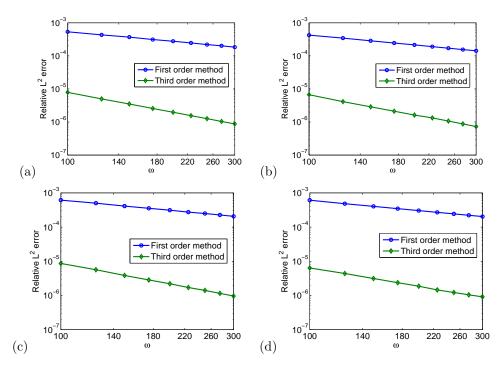
FIG. 5.4. *Relative L^2 error. (a): point wave. (b): plane wave with $\theta_0 = \frac{\pi}{2} + \frac{\pi}{4\lceil \sqrt{\omega} \rceil}$. (c): cylindrical wave. (d): variable velocity problem with $a = 40$.*
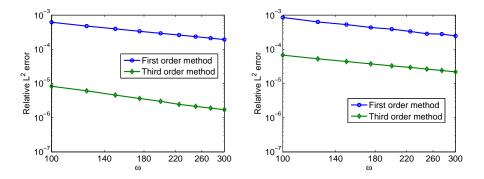

FIG. 5.5. *Convergence rate for smaller a. Left: $a = 20$. Right: $a = 5$.*

resolve one wave length.

The theoretical investigation in the last section reveals that setting $M = O(\omega^{1/2})$ is sufficient to ensure the asymptotic validity of the proposed Gaussian beam approach. This assertion is also supported by the numerical evidences shown in figure 5.3, where we plot the numerical errors by fixing $N_p = 4\lceil \omega \rceil$. The error saturates for large enough M. The location of the breakpoints is not as clear as that shown in figure 5.2 (notice the difference of scales), but one can still observe that the errors for $M \geq 4\lceil \omega^{1/2} \rceil$ always remain stable.

Next, we make a numerical investigation of the asymptotic convergence rate of the first and third Gaussian beam approaches. Here and hereafter, we set $N_p = 4\lceil\omega\rceil$ and $M = 4\lceil\sqrt{\omega}\rceil$. In figure 5.4 we show the relative L^2 errors on the whole domain of definition for four different problems. In the frequency range $[100,300]$, the convergence rates of the first and third order approaches are 0.973 and 1.998 for the point wave, 0.999 and 1.993 for the plane wave with $\theta_0 = \frac{\pi}{2} + \frac{\pi}{4\lceil\sqrt{\omega}\rceil}$, 0.996 and 2.000 for the cylindrical wave, 0.998 and 1.801 for the variable velocity problem with $a = 40$. Except for the variable velocity problem, the convergence rates are almost optimal; see Theorem 4.5 and Remark 4.7. The rate reduction in the case of variable velocity is probably because the frequency ω does not fall into the asymptotic regime yet. In fact, if we enlarge the velocity fluctuation by decreasing the parameter a to 20 and 5, the convergence rate of the third order approach deteriorates further to 1.465 and 1.015, see figure 5.5. In all cases, however, the error of the third order approach is much smaller than the first order.

6. Domain decomposition technique

The theoretical analysis conducted in section 4 and the numerical tests performed in the last section demonstrate that if the wave field mainly consists of the propagating plane waves, our Gaussian beam approach could present asymptotically convergent numerical solutions; see Remark 4.7. However, for a general boundary value problem, this condition cannot be taken for granted. For example, let us consider the constant velocity problem with $c(\mathbf{r}) \equiv 1$. In this case, the Helmholtz equation admits the wave solution

$$u = \exp\{-\alpha(x+0.5) + i\beta y\}, \tag{6.1}$$

where $\alpha > 0$ and $\beta = \sqrt{\omega^2 + \alpha^2}$. This solution propagates in the positive y-direction, but decays in the positive x-direction with a factor α. If $\alpha = O(\omega^{1/2})$, Theorem 3.1 and Hypothesis 3.4 ensure an element of the beam space Φ (see (4.5)) which approximates (6.1) within an error of $O(\omega^{-\gamma_N})$. Thus one could expect that in this case the Gaussian beam approach would still work. This is indeed true and verified by the numerical evidence shown in figure 6.1, where the relative L^2 error on the whole domain of definition for $\alpha = \omega^{1/2}$ is plotted. However, if the decay of the wave solution becomes more severe, say $\alpha = O(\omega)$, things change dramatically. In figure 6.2 we plot the computed solutions for $\alpha = \omega/2$ with $\omega = 100$ and $\omega = 200$. We see that the computed solutions are obviously wrong since the maximum modulus should be 1. The reason for this is simple: since the evanescent wave solution produces a boundary layer of thickness $O(\omega^{-1})$, it could not be approximated by any summation of Gaussian beams with characteristic width $\omega^{-1/2}$.

The above numerical tests naturally suggest application of the domain decomposition technique for handling the Helmholtz equation (1.1) with a general boundary data. The whole convex domain Ω is decomposed into two parts: a boundary layer region Ω_b and its complementary interior convex region Ω_i. The interface of Ω_b and Ω_i is denoted by Γ_i. The thickness of Ω_b is set as $O(\omega^{-1/2})$, so that the boundary evanescent waves (stronger than $O(\omega^{1/2})$) decays to zero sufficiently at Γ_i. The solution in Ω_b is then computed by a domain-based discretization method such as eighth-order FEM used in the following numerical tests, and the solution in Ω_i is computed by the Gaussian beam approach. Schwarz iterations are then performed by applying suitable transmission boundary conditions. In this paper, we simply employ the zeroth order absorbing transmission boundary condition together with the under-
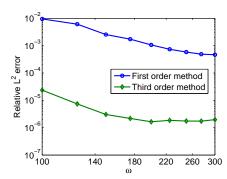
FIG. 6.1. *Relative L^2 errors for the weakly damping plane waves. $\alpha = \omega^{1/2}$. Third order method is used.*
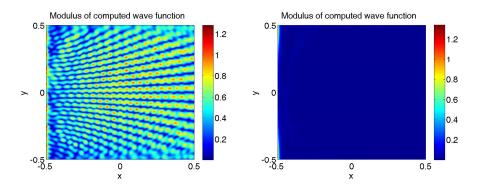

FIG. 6.2. *Modulus of computed wave function. The third order Gaussian beam approach is used. Left: $\omega = 100$. Right: $\omega = 200$.*

relaxation technique [8]. More sophisticated transmission boundary conditions could also be considered, see [24, 19], but they will not be pursued here.

We reconsider the numerical example for $\alpha = \omega/2$ by the Gaussian beam approach together with the domain decomposition technique. The boundary layer region Ω_b is discretized with eighth-order quadrilateral finite elements. The thickness of Ω_b is set as $0.125 > 100^{-1/2}$, and the mesh size is set as 0.025. Figure 6.3 illustrates the computed solution and the error function for $\omega = 100$, and figure 6.4 for $\omega = 200$. The relative infinity norm is less than 0.09 percent for $\omega = 100$, and 0.02 percent for $\omega = 200$. Moreover, in the case of $\omega = 200$, one could notice that the error mainly comes from the finite element discretization of the boundary layer region Ω_b. It is thus promising to further reduce the overall error by using a more refined mesh in Ω_b.

As a final numerical test, we compute the solution of the constant velocity Helmholtz equation (1.1) for $\omega = 100$ equipped with the boundary condition

$$(\partial_\nu - i\omega)u(\mathbf{r}) = -i\omega.$$

The thickness of boundary layer region and the element size are as before. The numerical solution by the third order Gaussian beam approach together with the domain decomposition technique is plotted in the left of figure 6.5. To evaluate the quality of numerical solution, we need a reference solution. Note that the function
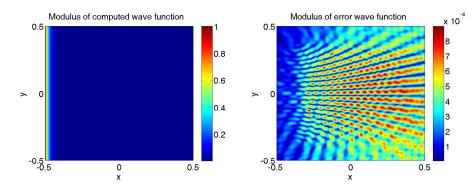
FIG. 6.3. $\omega = 100$. *Third order Gaussian beam approach together with the domain decomposition technique. Left: Modulus of the computed wave field. Right: Modulus of the error wave function.*

FIG. 6.4. $\omega = 200$. *Third order Gaussian beam approach together with the domain decomposition technique. Left: Modulus of the computed wave field. Right: Modulus of the error wave function.*
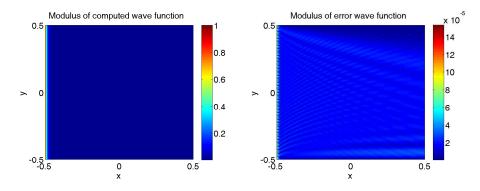

FIG. 6.5. $\omega = 100$. *Third order Gaussian beam approach together with the domain decomposition technique. Left: Modulus of the computed wave field. Right: Modulus of the error wave function.*
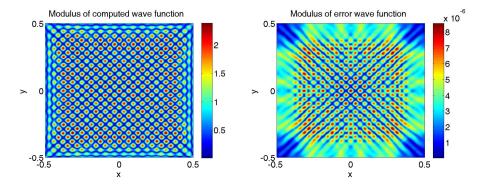
$\tilde{u} = u - 1$ satisfies

$$\Delta \tilde{u} + \omega^2 \tilde{u} = -\omega^2, \ \mathbf{r} \in \Omega = [-0.5, 0.5]^2,$$
$$\partial_\nu \tilde{u} - i\omega \tilde{u} = 0, \ \mathbf{r} \in \partial\Omega.$$

We compute $\tilde{u}$ on the whole domain of definition Ω by the same finite element method

with the same mesh size. The resulting linear system is solved by the characteristic decomposition method. This is possible since both the geometry Ω and the governing equation are separable. In the right of figure 6.5, we plot the error between the solution by the Gaussian beam method and the reference solution $\tilde{u}+1$. The maximum error is less than 9×10^{-6}.

7. Conclusion and discussion

We proposed an asymptotic solver called the Gaussian beam approach for the high frequency Helmholtz boundary value problem. The basic idea is to seek the approximate solution by the least squares algorithm in a linear space spanned by a set of Gaussian beams. Gaussian beams are asymptotic solutions of the linear wave equation. For the Helmholtz equation, we explicitly deduced the formulations for the first three order Gaussian beams. In the case of constant velocity, these formulations were integrated out analytically. The key ingredient of the proposed approach is the construction of a beam space which has a good approximating potential for traveling waves. We gave a theoretical analysis on the accuracy of Gaussian beam approximations for plane waves. The accuracy of the numerical solution obtained by our approach was also investigated based on some a priori asymptotic assumptions. Numerical tests demonstrated the effectiveness of the proposed method, and showed the superiority of the third order method to the first order method.

If the exact wave solution contains some strongly evanescent wave modes decaying stronger than $O(\omega^{1/2})$, the direct implementation of the Gaussian beam approach might fail. We proposed the domain decomposition technique to handle this situation. The domain of definition is decomposed into a boundary layer region of thickness $O(\omega^{-1/2})$, and its complementary interior traveling wave region. The boundary layer region could be discretized by any domain-based method, and the traveling wave region is solved by the Gaussian beam approach. Schwarz iterations should then be performed with any efficient transmission boundary condition developed by the domain decomposition technique community. We remark that the convexity assumption of the domain of definition is not essential to our approach. Combined with the domain decomposition technique, the Gaussian beam approach could in fact handle various high frequency linear wave problems in more complicated geometries.

A remarkable advantage of the Gaussian beam approach lies in the fact that it imposes a weak restriction on the computer memory for large frequency ω. The Gaussian beam approach involves only $O(\omega)$ unknowns, much less than any domain-based discretization method. Even compared with the integral equation method which is generally only applicable for the constant velocity problems, the Gaussian beam approach would result in a much sparser linear algebraic problem. Since each beam has a width of $O(\omega^{-1/2})$, the operator $\mathcal{B}$ involves only $O(\omega^{3/2})$ nonzero elements in the discrete level. Thus each iteration needs only $O(\omega^{3/2})$ operations. However, we should confess that the iterative solver CGLS for the normal equation (4.13), though working quite well, is not very efficient to this moment. The numerical experiements showed that the iteration number typically increases with the frequency ω. We anticipate that a suitable preconditioning technique would greatly speed up the iteration process. This is very important to make the Gaussian beam approach more attractive, since for this moment the most sophisticated solver based on the domain discretization for the two-dimensional Helmholtz equation needs only $O(\omega^2)$ operations. This issue is currently under investigation.

Appendix A. In the constant velocity case $c(\mathbf{r}) \equiv 1$, the ODE system (2.16)–(2.23) can be integrated out analytically. The parameters involved in the first three Gaussian beam solutions under the initial conditions (2.27) are listed in the following

$$\phi_2 = \frac{a_2}{1 + a_2 t},$$

$$\phi_3 = \frac{a_3}{(1 + a_2 t)^3},$$

$$\phi_4 = \frac{a_4}{(1 + a_2 t)^4} - \frac{3t\left(a_2^5 t + a_2^4 + a_3^2\right)}{(1 + a_2 t)^5},$$

$$A_{00} = \frac{a_{00}}{\sqrt{1 + a_2 t}},$$

$$A_{01} = \frac{a_{01}}{(1 + a_2 t)^{\frac{3}{2}}} - \frac{a_{00} a_3 t}{2(1 + a_2 t)^{\frac{5}{2}}},$$

$$A_{02} = -\frac{1}{4}(1 + a_2 t)^{-\frac{9}{2}}\left(6 a_{00} a_2^4 t^2 + 6 a_{00} a_2^3 t - 4 a_{02} a_2^2 t^2 + 8 a_{01} a_2 a_3 t^2 \right.$$

$$\left. + 2 a_{00} a_4 a_2 t^2 - 8 a_{02} a_2 t - 5 a_{00} a_3^2 t^2 + 8 a_{01} a_3 t + 2 a_{00} a_4 t - 4 a_{02}\right),$$

$$A_1 = \frac{a_1}{\sqrt{1 + a_2 t}} - \frac{t}{24}(1 + a_2 t)^{-\frac{7}{2}}\left(9 a_{00} a_2^3 t + 12 a_{02} a_2^2 t^2 + 9 a_{00} a_2^2\right.$$

$$\left. - 12 a_{01} a_2 a_3 t^2 - 3 a_{00} a_4 a_2 t^2 + 24 a_{02} a_2 t + 5 a_{00} a_3^2 t^2 - 12 a_{01} a_3 t - 3 a_{00} a_4 t + 12 a_{02}\right).$$

Appendix B. Given $\gamma > 0$ and $c < 3/2$, the following holds:

$$\sum_{k=1}^{\infty} \exp(-\gamma(k-c)^2)) \leq \frac{\exp(-\gamma(1-c)^2))}{1 - \exp(-\gamma(3-2c))}.$$

Moreover, if $|c| \leq 1/4$, we have

$$\sum_{k=1}^{\infty} \exp(-\gamma(k-c)^2)) \leq \frac{\exp(-9\gamma/16))}{1 - \exp(-5\gamma/2)}.$$

Proof. Since

$$\sum_{k=1}^{\infty} \exp(-\gamma(k-c)^2)) = \exp(-\gamma(1-c)^2)) + \sum_{k=1}^{\infty} \exp(-\gamma(k+1-c)^2))$$

$$= \exp(-\gamma(1-c)^2)) + \sum_{k=1}^{\infty} \exp(-\gamma(k-c)^2)) \exp(-\gamma(2(k-c)+1))$$

$$\leq \exp(-\gamma(1-c)^2)) + \exp(-\gamma(3-2c)) \sum_{k=1}^{\infty} \exp(-\gamma(k-c)^2)),$$

thus

$$\sum_{k=1}^{\infty} \exp(-\gamma(k-c)^2)) \leq \frac{\exp(-\gamma(1-c)^2))}{1-\exp(-\gamma(3-2c))}.$$

In the interval $[-1/4, 1/4]$, the right hand side is an increasing function of c. Taking $c = 1/4$ we get the second assertion. □

Appendix C. For any $z \in \mathbb{C}$ satisfying $|z| \leq 1$, we have

$$|\exp(z) - 1| \leq 1.7183|z|.$$

Proof.   Since $\exp(z)$ is an entire function, we have

$$|\exp(z) - 1| = \left| \sum_{k=1}^{\infty} \frac{z^k}{k!} \right| \leq \sum_{k=1}^{\infty} \frac{|z|^k}{k!} = \exp(|z|) - 1 \leq (\exp(1) - 1)|z| < 1.7183|z|.$$

This ends the proof. □

REFERENCES

[1] I.M. Babuska and S.A. Sauter, *Is the pollution effect of the FEM avoidable for the Helmholtz equation considering high wave numbers*, SIAM Review, 42(3), 451–484, 2000.

[2] A. Björk, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, PA, 1996.

[3] V. Cerveny, *Expansion of a plane wave into Gaussian beams*, Studia geoph. et geod., 26, 120–131, 1982.

[4] V. Cerveny, *Seismic Ray Theory*, Cambridge Univ. Press, Cambridge, 2001.

[5] V. Cerveny, M.M. Popov and I. Psencik, *Computation of wave fields in inhomogeneous media – Gaussian beam approach*, Geophys. J. R. Astr. Soc., 70, 109–128, 1982.

[6] G. Cincotti, F. Gori, M. Santarsiero, F. Frezza, F. Furnò and G. Schettini, *Plane wave expansion of cylindrical functions*, Optics Communications, 95, 192–198, 1993.

[7] P. Cummings and X.B. Feng, *Sharp regularity coefficient estimates for complex-valued acoustic and elastic Helmholtz equations*, Math. Models Meth. Appl. Sci., 16(1), 139–160, 2006.

[8] B. Despres, *Domain decomposition method and the Helmholtz equation. II*, in second Int. Conf. Math. Numerical Aspects of Wave Propagation, Philadelphia, PA, SIAM, 197–206, 1993.

[9] B. Engquist and O. Runborg, *Computational high frequency wave propagation*, Acta Numerica, 12, 181–266, 2003.

[10] C.W. Groetsch, *Inverse Problems in the Mathematical Sciences*, Vieweg Verlag, Wiesbaden, 1993.

[11] S. Gutman and A.G. Ramm, *Numerical implementation of the MRC method for obstacle scattering problems*, J. Phys. A: Math. Gen., 35, 8065–8074, 2002.

[12] P.C. Hansen, *Numerical tools for analysis and solution of Fredholm integral equations of the first kind*, Inverse Problems, 8, 849–872, 1992.

[13] C.M.R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Stand., 49, 409–436, 1952.

[14] B. Hoffmann, *Regularization for Applied Inverse and Ill-posed Problems*, Teubner, Leipzig, 1986.

[15] J. B. Keller and R. Lewis, *Asymptotic Methods for Partial Differential Equations: the Reduced Wave Equation and Maxwell's Equation*, in Surveys in Applied Mathematics, J.B. Keller, D.McLaughlin and G. Papanicolaou (eds.), Plenum Press, New York, 1995.

[16] L. Klimes, *Discretization error for the superposition of Gaussian beams*, Geophys. J. R. astr. Soc., 86, 531–551, 1986.

[17] L. Klimes, *Second-order and higher-order perturbations of travel time in isotropic and anisotropic media*, Studia geop. et geod., 46, 213–248, 2002.

[18] L. Klimes, *Spatial derivatives and perturbation derivatives of amplitude in isotropic and anisotropic media*, Studia geop. et geod., 50, 417–430, 2006.

[19] F. Magoules, P. Ivanyi and B.H.V. Topping, *Non-overlapping schwarz methods with optimized transmission conditions for the Helmholtz equation*, Comput. Meth. Appl. Mech. Engrg., 193, 4797–4818, 2004.

[20] M. Motamed and O. Runborg, *Taylor expansion and discretization errors in Gaussian beam superposition*, to appear in Wave Motion, 2010.

[21] M.M. Popov, *A new method of computation of wave fields using Gaussian beams*, Wave Motion, 4, 85–97, 1982.

[22] Y. Saad, *Iterative Methods for Sparse Linear Systems*, PWS, Boston, 1996.

[23] N.M. Tanushev, J. Qian and J.V. Ralston, *Mountain waves and Gaussian beams*, SIAM Multiscale Modeling and Simulation, 6(2), 688–709, 2007.

[24] A. Toselli, *Overlapping methods with perfectly matched layers for the solution of the Helmholtz equation*, in Proceedings of the 11th Intl. Conf. on Domain Decomposition Methods, Domain Decomposition Press, 551–557, 1999.

[25] N. Weck, *Approximation by Herglotz wave functions*, Math. Meth. Appl. Sci., 27, 155–162, 2004.