FROZEN GAUSSIAN APPROXIMATION FOR HIGH FREQUENCY WAVE PROPAGATION[*]

JIANFENG LU[†] AND XU YANG[‡]

Abstract. We propose the frozen Gaussian approximation for computation of high frequency wave propagation. This method approximates the solution to the wave equation by an integral representation. It provides a highly efficient computational tool based on the asymptotic analysis on the phase plane. Compared to geometric optics, it provides a valid solution around caustics. Compared to the Gaussian beam method, it overcomes the drawback of beam spreading. We give several numerical examples to verify that the frozen Gaussian approximation performs well in the presence of caustics and when the Gaussian beam spreads. Moreover, it is observed numerically that the frozen Gaussian approximation exhibits better accuracy than the Gaussian beam method.

Key words. High frequency wave propagation, Frozen Gaussian approximation, wave equation, phase plane analysis, asymptotics.

AMS subject classifications. 81Q20, 35L05, 65M99.

1. Introduction

We are interested in developing efficient numerical methods for high frequency wave propagation. For simplicity and clarity we take the following linear scalar wave equation to present the idea:

$$\partial_t^2 u - c^2(\boldsymbol{x})\Delta u = 0, \qquad \boldsymbol{x} \in \mathbb{R}^d, \tag{1.1}$$

with WKB initial conditions

$$\begin{cases} u_0(\boldsymbol{x}) = A_0(\boldsymbol{x})e^{\frac{i}{\varepsilon}S_0(\boldsymbol{x})}, \\ \partial_t u_0(\boldsymbol{x}) = \frac{1}{\varepsilon}B_0(\boldsymbol{x})e^{\frac{i}{\varepsilon}S_0(\boldsymbol{x})}, \end{cases} \tag{1.2}$$

where u is the wave field, d is the dimensionality and $i = \sqrt{-1}$ is the imaginary unit. We assume that the local wave speed $c(\boldsymbol{x})$ is a smooth function. The small parameter $\varepsilon \ll 1$ characterizes the high frequency nature of the wave. The proposed method can be generalized to other types of wave equations [17].

Numerical computation of high frequency wave propagation is an important problem arising in many applications, such as electromagnetic radiation and scattering, and seismic and acoustic waves traveling, just to name a few. It is a two-scale problem. The *large* length scale comes from the characteristic size of the medium, while the *small* length scale is the wavelength. The disparity between the two length scales

[†]Department of Mathematics, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, (jianfeng@cims.nyu.edu).

[‡]Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544, (xuyang@math.princeton.edu).

makes direct numerical computations extremely hard. In order to achieve accurate results, the mesh size has to be chosen comparable to the wavelength or even smaller. On the other hand, the domain size is large so that a huge number of grid points are needed.

In order to compute efficiently high frequency wave propagation, algorithms based on asymptotic analysis have been developed. One of the most famous examples is geometric optics. In the method, it is assumed that the solution has the form

$$u(t,\boldsymbol{x}) = A(t,\boldsymbol{x})e^{\imath S(t,\boldsymbol{x})/\varepsilon}. \tag{1.3}$$

To the leading order, the phase function $S(t,\boldsymbol{x})$ satisfies the eikonal equation,

$$|\partial_t S|^2 - c^2(\boldsymbol{x})|\nabla_{\boldsymbol{x}}S|^2 = 0, \tag{1.4}$$

and the amplitude $A(t,\boldsymbol{x})$ satisfies the transport equation,

$$\partial_t A - c^2(\boldsymbol{x})\frac{\nabla_{\boldsymbol{x}}S}{\partial_t S}\cdot\nabla_{\boldsymbol{x}}A + \frac{\left(\partial_t^2 S - c^2(\boldsymbol{x})\Delta S\right)}{2\partial_t S}A = 0.$$

The merit of geometric optics is that it only solves the macroscopic quantities $S(t,\boldsymbol{x})$ and $A(t,\boldsymbol{x})$ which are ε-independent. Computational methods based on the geometric optics are reviewed in [3, 26].

However, since the eikonal Equation (1.4) is of Hamilton-Jacobi type, the solution of (1.4) becomes singular after the formation of caustics. At caustics, the approximate solution of geometric optics is invalid since the amplitude $A(t,\boldsymbol{x})$ blows up. To overcome this problem, Popov introduced the Gaussian beam method in [19]. The single beam solution of the Gaussian beam method has a similar form to geometric optics,

$$u(t,\boldsymbol{x}) = A(t,\boldsymbol{y})e^{\imath \tilde{S}(t,\boldsymbol{x},\boldsymbol{y})/\varepsilon}.$$

The difference lies in that the Gaussian beam method uses a *complex* phase function,

$$\tilde{S}(t,\boldsymbol{x},\boldsymbol{y}) = S(t,\boldsymbol{y}) + \boldsymbol{p}(t,\boldsymbol{y})\cdot(\boldsymbol{x}-\boldsymbol{y}) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{y})\cdot M(t,\boldsymbol{y})(\boldsymbol{x}-\boldsymbol{y}), \tag{1.5}$$

where $S\in\mathbb{R}, \boldsymbol{p}\in\mathbb{R}^d, M\in\mathbb{C}^{d\times d}$. The imaginary part of M is chosen to be positive definite so that the solution decays exponentially away from $\boldsymbol{x}=\boldsymbol{y}$, where $\boldsymbol{y}$ is called the beam center. This makes the solution a Gaussian function, and hence the method was named the Gaussian beam method. If the initial wave is not in a form of a single beam, one can approximate it by using a number of Gaussian beams. The validity of this construction at caustics was analyzed by Ralston in [24]. Recently, there have been a series of numerical studies including both the Lagrangian type [18, 22, 23, 28, 29] and the Eulerian type [9, 8, 10, 11, 14, 15, 20, 21].

The construction of Gaussian beam approximation is based on the truncation of the Taylor expansion of $\tilde{S}$ around the beam center $\boldsymbol{y}$ up to the quadratic term, hence it loses accuracy when the width of the beam becomes large, *i.e.*, when the imaginary part of $M(t,\boldsymbol{y})$ in (1.5) becomes small so that the Gaussian function is no longer localized. This happens, for example, when the solution of the wave equation spreads (the opposite situation of forming caustics). This is a severe problem in general, as shown by the examples in Section 4. One could overcome the problem of spreading of beams by doing reinitialization once in a while; see [22, 23]. This increases the computational complexity, especially when beams spread quickly.

Therefore a method working in both scenarios (spreading and caustics) is required. The main idea of the method proposed in the current work is to use Gaussian functions with fixed widths, instead of using those that might spread over time, to approximate the wave solution. That is why this type of method is called frozen Gaussian approximation (FGA). Despite its superficial similarity with the Gaussian beam method (GBM), it is different at a fundamental level. FGA is based on phase plane analysis, while GBM is based on the asymptotic solution to a wave equation with Gaussian initial data. In FGA, the solution to the wave equation is approximated by a superposition of Gaussian functions living in the phase space, and each function is *not* necessarily an asymptotic solution, while GBM uses Gaussian functions (named as beams) in the physical space, with each individual beam being an asymptotic solution to the wave equation. The main advantage of FGA over GBM is that the problem of beam spreading no longer exists.[1] Besides, numerically we observe that FGA has better accuracy than GBM when keeping the same order of terms in asymptotic series. On the other hand, the solution given by FGA is asymptotically accurate around caustics where geometric optics breaks down.

Our work is motivated by the chemistry literature on the propagation of time dependent Schrödinger equations, where the spreading of solution is a common phenomenon, for example, in the dynamics of a free electron. In [5], Heller introduced frozen Gaussian wavepackets to deal with this issue, but it only worked for a short time propagation of order $\mathcal{O}(\hbar)$ where $\hbar$ is the Planck constant. To make it valid for longer time of order $\mathcal{O}(1)$, Herman and Kluk proposed in [6] to change the weight of Gaussian packets by adding so-called Herman-Kluk prefactor. Integral representation and higher order approximations were developed by Kay in [12] and [13]. Recently, the semiclassical approximation underlying the method was analyzed rigorously by Swart and Rousse in [27] and also by Robert in [25]. We generalize their ideas for propagation of high frequency waves, aiming at developing an efficient computational method. We decompose waves into several branches of propagation, and each of them is approximated using Gaussian functions on the phase plane. Their centers follow different Hamiltonian dynamics for different branches. Their weight functions, which are analogous to the Herman-Kluk prefactor, satisfy new evolution equations derived from asymptotic analysis.

The rest of paper is organized as follows. In Section 2, we state the formulations and numerical algorithm of the frozen Gaussian approximation. In Section 3, we provide asymptotic analysis to justify the formulations introduced in Section 2. The numerical examples are given in Section 4 to verify the accuracy and to compare the frozen Gaussian approximation (FGA) with the Gaussian beam method (GBM). In Section 5, we discuss the efficiency of FGA in comparison with GBM and higher order GBM, with some comments on the phenomenon of error cancellation, and we give some conclusive remarks in the end.

2. Formulation and algorithm

In this section we present the basic formulation and the main algorithm of the frozen Gaussian approximation (FGA), and leave the derivation to the next section.

---

[1]Divergence is still an issue for the Lagrangian approach; one needs to work in the Eulerian framework to completely solve the problem, which is considered in [17].

2.1. Formulation. FGA approximates the solution to the wave Equation (1.1) by the integral representation

$$
\begin{aligned}
u^{\mathrm{FGA}}(t,\boldsymbol{x}) = {} & \frac{1}{(2\pi\varepsilon)^{3d/2}} \int_{\mathbb{R}^{3d}} a_+(t,\boldsymbol{q},\boldsymbol{p}) e^{\frac{i}{\varepsilon}\Phi_+(t,\boldsymbol{x},\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})} u_{+,0}(\boldsymbol{y}) \,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q} \\
& + \frac{1}{(2\pi\varepsilon)^{3d/2}} \int_{\mathbb{R}^{3d}} a_-(t,\boldsymbol{q},\boldsymbol{p}) e^{\frac{i}{\varepsilon}\Phi_-(t,\boldsymbol{x},\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})} u_{-,0}(\boldsymbol{y}) \,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q},
\end{aligned}
\tag{2.1}
$$

where $u_{\pm,0}$ are determined by the initial value

$$
u_{\pm,0}(\boldsymbol{x}) = A_\pm(\boldsymbol{x}) e^{\frac{i}{\varepsilon}S_0(\boldsymbol{x})},
\tag{2.2}
$$

with

$$
A_\pm(\boldsymbol{x}) = \frac{1}{2}\left(A_0(\boldsymbol{x}) \pm \frac{i B_0(\boldsymbol{x})}{c(\boldsymbol{x})|\partial_{\boldsymbol{x}} S_0(\boldsymbol{x})|}\right).
$$

The Equation (2.1) implies that the solution consists of two branches ("$\pm$").

In (2.1), $\Phi_\pm$ are the phase functions given by

$$
\begin{aligned}
\Phi_\pm(t,\boldsymbol{x},\boldsymbol{y},\boldsymbol{q},\boldsymbol{p}) = {} & \boldsymbol{P}_\pm(t,\boldsymbol{q},\boldsymbol{p}) \cdot (\boldsymbol{x} - \boldsymbol{Q}_\pm(t,\boldsymbol{q},\boldsymbol{p})) - \boldsymbol{p}\cdot(\boldsymbol{y}-\boldsymbol{q}) \\
& + \frac{i}{2}|\boldsymbol{x} - \boldsymbol{Q}_\pm(t,\boldsymbol{q},\boldsymbol{p})|^2 + \frac{i}{2}|\boldsymbol{y}-\boldsymbol{q}|^2.
\end{aligned}
\tag{2.3}
$$

Given $\boldsymbol{q}$ and $\boldsymbol{p}$ as parameters, the evolution of $\boldsymbol{Q}_\pm$ and $\boldsymbol{P}_\pm$ are given by the equation of motion corresponding to the Hamiltonian $H_\pm = \pm c(\boldsymbol{Q}_\pm)|\boldsymbol{P}_\pm|$,

$$
\begin{cases}
\dfrac{\mathrm{d}\boldsymbol{Q}_\pm}{\mathrm{d}t} = \partial_{\boldsymbol{P}_\pm} H_\pm = \pm c\dfrac{\boldsymbol{P}_\pm}{|\boldsymbol{P}_\pm|}, \\[2mm]
\dfrac{\mathrm{d}\boldsymbol{P}_\pm}{\mathrm{d}t} = -\partial_{\boldsymbol{Q}_\pm} H_\pm = \mp \partial_{\boldsymbol{Q}_\pm} c\,|\boldsymbol{P}_\pm|,
\end{cases}
\tag{2.4}
$$

with the initial conditions $\boldsymbol{Q}_\pm(0,\boldsymbol{q},\boldsymbol{p}) = \boldsymbol{q}$ and $\boldsymbol{P}_\pm(0,\boldsymbol{q},\boldsymbol{p}) = \boldsymbol{p}$. The evolution equation of $a_\pm$ is given by

$$
\begin{aligned}
\frac{\mathrm{d}a_\pm}{\mathrm{d}t} = {} & \pm \frac{a_\pm}{2}\left(\frac{\boldsymbol{P}_\pm}{|\boldsymbol{P}_\pm|}\cdot \partial_{\boldsymbol{Q}_\pm} c - \frac{(d-1)i}{|\boldsymbol{P}_\pm|}c\right) \\
& \pm \frac{a_\pm}{2}\mathrm{tr}\left(Z_\pm^{-1}\partial_{\boldsymbol{z}}\boldsymbol{Q}_\pm\left(2\frac{\boldsymbol{P}_\pm}{|\boldsymbol{P}_\pm|}\otimes \partial_{\boldsymbol{Q}_\pm} c\right.\right. \\
& \left.\left. - \frac{ic}{|\boldsymbol{P}_\pm|}\left(\frac{\boldsymbol{P}_\pm\otimes \boldsymbol{P}_\pm}{|\boldsymbol{P}_\pm|^2} - I\right) - i|\boldsymbol{P}_\pm|\partial_{\boldsymbol{Q}_\pm}^2 c\right)\right),
\end{aligned}
\tag{2.5}
$$

with the initial condition

$$
a_\pm(0,\boldsymbol{q},\boldsymbol{p}) = 2^{d/2}.
$$

In (2.5), $\boldsymbol{P}_\pm$ and $\boldsymbol{Q}_\pm$ are evaluated at $(t,\boldsymbol{q},\boldsymbol{p})$, c and $\partial_{\boldsymbol{Q}_\pm} c$ are evaluated at $\boldsymbol{Q}_\pm$, I is the identity matrix, and we have introduced short hand notations

$$
\partial_{\boldsymbol{z}} = \partial_{\boldsymbol{q}} - i\partial_{\boldsymbol{p}}, \qquad Z_\pm = \partial_{\boldsymbol{z}}(\boldsymbol{Q}_\pm + i\boldsymbol{P}_\pm).
\tag{2.6}
$$

The evolution of the weight $a_\pm$ is analogous to the Herman-Kluk prefactor [6].

REMARK 2.1.
    1. The Equation (2.5) can be reformulated as

$$\frac{\mathrm{d}a_\pm}{\mathrm{d}t} = \pm a_\pm \frac{\boldsymbol{P}_\pm}{|\boldsymbol{P}_\pm|} \cdot \partial_{\boldsymbol{Q}_\pm} c + \frac{a_\pm}{2} \operatorname{tr}\left(Z_\pm^{-1} \frac{\mathrm{d}Z_\pm}{\mathrm{d}t} \right). \tag{2.7}$$

When c is constant, (2.7) has an analytical solution $a_\pm = (\det Z_\pm)^{1/2}$ with the branch of square root determined continuously in time by the initial value.
    2. $\partial_{\boldsymbol{z}} \boldsymbol{Q}_\pm$ and $\partial_{\boldsymbol{z}} \boldsymbol{P}_\pm$ satisfy the following evolution equations:

$$\frac{\mathrm{d}(\partial_{\boldsymbol{z}} \boldsymbol{Q}_\pm)}{\mathrm{d}t} = \pm \partial_{\boldsymbol{z}} \boldsymbol{Q}_\pm \frac{\partial_{\boldsymbol{Q}_\pm} c \otimes \boldsymbol{P}_\pm}{|\boldsymbol{P}_\pm|} \pm c \partial_{\boldsymbol{z}} \boldsymbol{P}_\pm \left(\frac{I}{|\boldsymbol{P}_\pm|} - \frac{\boldsymbol{P}_\pm \otimes \boldsymbol{P}_\pm}{|\boldsymbol{P}_\pm|^3} \right), \tag{2.8}$$

$$\frac{\mathrm{d}(\partial_{\boldsymbol{z}} \boldsymbol{P}_\pm)}{\mathrm{d}t} = \mp \partial_{\boldsymbol{z}} \boldsymbol{Q}_\pm \partial_{\boldsymbol{Q}_\pm}^2 c |\boldsymbol{P}_\pm| \mp \partial_{\boldsymbol{z}} \boldsymbol{P}_\pm \frac{\boldsymbol{P}_\pm \otimes \partial_{\boldsymbol{Q}_\pm} c}{|\boldsymbol{P}_\pm|}. \tag{2.9}$$

One can solve (2.8)-(2.9) to get $\partial_{\boldsymbol{z}} \boldsymbol{Q}_\pm$ and $\partial_{\boldsymbol{z}} \boldsymbol{P}_\pm$ in (2.5). This increases the computational cost, but avoids the errors of using divided differences to approximate derivatives.
    Notice that (2.1) can be rewritten as

$$\begin{aligned}
u^{\mathrm{FGA}}(t,\boldsymbol{x}) = &\int_{\mathbb{R}^{2d}} \frac{a_+}{(2\pi\varepsilon)^{3d/2}} \psi_+ e^{\frac{i}{\varepsilon} \boldsymbol{P}_+ \cdot (\boldsymbol{x} - \boldsymbol{Q}_+) - \frac{1}{2\varepsilon}|\boldsymbol{x} - \boldsymbol{Q}_+|^2} \, \mathrm{d}\boldsymbol{p} \, \mathrm{d}\boldsymbol{q} \\
&+ \int_{\mathbb{R}^{2d}} \frac{a_-}{(2\pi\varepsilon)^{3d/2}} \psi_- e^{\frac{i}{\varepsilon} \boldsymbol{P}_- \cdot (\boldsymbol{x} - \boldsymbol{Q}_-) - \frac{1}{2\varepsilon}|\boldsymbol{x} - \boldsymbol{Q}_-|^2} \, \mathrm{d}\boldsymbol{p} \, \mathrm{d}\boldsymbol{q},
\end{aligned} \tag{2.10}$$

where

$$\psi_\pm(\boldsymbol{q},\boldsymbol{p}) = \int_{\mathbb{R}^d} u_{\pm,0}(\boldsymbol{y}) e^{-\frac{i}{\varepsilon} \boldsymbol{p} \cdot (\boldsymbol{y} - \boldsymbol{q}) - \frac{1}{2\varepsilon}|\boldsymbol{y} - \boldsymbol{q}|^2} \, \mathrm{d}\boldsymbol{y}. \tag{2.11}$$

Therefore, the method first decomposes the initial wave into several Gaussian functions in phase space, and then propagates the center of each function along the characteristic lines while keeping the width of the Gaussian fixed. This vividly explains the name frozen Gaussian approximation of this method.
    The formulation above gives the leading order frozen Gaussian approximation with an error of $\mathcal{O}(\varepsilon)$. It is not hard to obtain higher order approximations by the asymptotics presented in Section 3. We will focus mainly on the leading order approximation in this paper and leave the higher order corrections and rigorous numerical analysis to future works.

    2.2. Algorithm.   We first give a description of the overall algorithm. To construct the frozen Gaussian approximation on a mesh of $\boldsymbol{x}$, one needs to compute the integral (2.10) numerically with a mesh of $(\boldsymbol{q},\boldsymbol{p})$. This will relate to the numerical computation of (2.11) with a mesh of $\boldsymbol{y}$. Hence three different meshes are needed in the algorithm. Moreover, the stationary phase approximation implies that $\psi_\pm$ in (2.11) is localized around the submanifold $\boldsymbol{p} = \nabla_{\boldsymbol{q}} S_0(\boldsymbol{q})$ on the phase plane for WKB initial conditions (1.2) when ε is small. This means we only need to put the mesh grids of $\boldsymbol{p}$ around $\nabla_{\boldsymbol{q}} S_0(\boldsymbol{q})$ initially to get a good approximation of the initial value. A one-dimensional example is given to illustrate this localization property of $\psi_\pm$ in Figure 2.1 (left). The associated mesh grids are shown in Figure 2.1 (right).
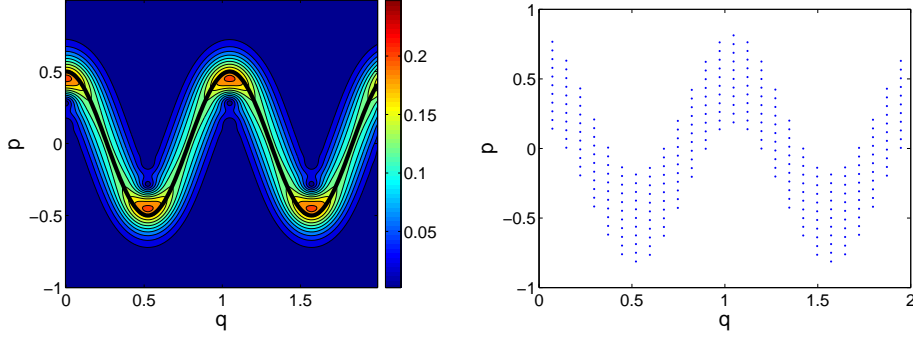   Next we describe in details all the meshes used in the algorithm.

FIG. 2.1.   *Left: an illustration of the localization of ψ_+ on (q,p) domain for $u_{+,0}(y) = \exp\left(\imath\frac{\sin(6y)}{12\varepsilon}\right)$, $\varepsilon = 1/128$; the black solid curve is $p = \cos(6q)/2$.   Right: the corresponding mesh grids of (q,p).*

1. Discrete mesh of $(\boldsymbol{q},\boldsymbol{p})$ for initializing $\boldsymbol{Q},\boldsymbol{P}$. Denote $\boldsymbol{\delta q} = (\delta q_1,\cdots,\delta q_d)$ and $\boldsymbol{\delta p} = (\delta p_1,\cdots,\delta p_d)$ as the mesh size. Suppose $\boldsymbol{q}^0 = (q_1^0,\cdots,q_d^0)$ is the starting point, then the mesh grids $\boldsymbol{q}^{\boldsymbol{k}}$, $\boldsymbol{k} = (k_1,\cdots,k_d)$, are defined as

$$\boldsymbol{q}^{\boldsymbol{k}} = \left(q_1^0 + (k_1-1)\delta q_1,\cdots,q_d^0 + (k_d-1)\delta q_d\right),$$

   where $k_j = 1,\cdots,N_q$ for each $j \in \{1,\cdots,d\}$.
   The mesh grids $\boldsymbol{p}^{\boldsymbol{k},\boldsymbol{\ell}}$, $\boldsymbol{\ell} = (\ell_1,\cdots,\ell_d)$, are defined associated with the mesh grids $\boldsymbol{q}^{\boldsymbol{k}}$,

$$\boldsymbol{p}^{\boldsymbol{k},\boldsymbol{\ell}} = \left(\partial_{q_1} S_0(\boldsymbol{q}^{\boldsymbol{k}}) + \ell_1\delta p_1,\cdots,\partial_{q_d} S_0(\boldsymbol{q}^{\boldsymbol{k}}) + \ell_d\delta p_d\right),$$

   where $\ell_j = -N_p,\cdots,N_p$ for each $j \in \{1,\cdots,d\}$.

2. Discrete mesh of $\boldsymbol{y}$ for evaluating $\psi_\pm$ in (2.11). $\boldsymbol{\delta y} = (\delta y_1,\cdots,\delta y_d)$ is the mesh size. Denote $\boldsymbol{y}^0 = (y_1^0,\cdots,y_d^0)$ as the starting point. The mesh grids $\boldsymbol{y}^{\boldsymbol{m}}$ are, for $\boldsymbol{m} = (m_1,\cdots,m_d)$,

$$\boldsymbol{y}^{\boldsymbol{m}} = \left(y_1^0 + (m_1-1)\delta y_1,\cdots,y_d^0 + (m_d-1)\delta y_d\right),$$

   where $m_j = 1,\cdots,N_y$ for each $j \in \{1,\cdots,d\}$.

3. Discrete mesh of $\boldsymbol{x}$ for reconstructing the final solution. $\boldsymbol{\delta x} = (\delta x_1,\cdots,\delta x_d)$ is the mesh size. Denote $\boldsymbol{x}^0 = (x_1^0,\cdots,x_d^0)$ as the starting point. The mesh grids $\boldsymbol{x}^{\boldsymbol{n}}$ are, for $\boldsymbol{n} = (n_1,\cdots,n_d)$,

$$\boldsymbol{x}^{\boldsymbol{n}} = \left(x_1^0 + (n_1-1)\delta x_1,\cdots,x_d^0 + (n_d-1)\delta x_d\right),$$

   where $n_j = 1,\cdots,N_x$ for each $j \in \{1,\cdots,d\}$.

With the preparation of the meshes, we introduce the algorithm as follows.

Step 1. Decompose the initial conditions (1.2) into two branches of waves according to (2.2).

Step 2. Compute the weight function $\psi_\pm$ by (2.11) for $(\boldsymbol{Q},\boldsymbol{P})$ initialized at $(\boldsymbol{q}^{\boldsymbol{k}},\boldsymbol{p}^{\boldsymbol{k},\boldsymbol{\ell}})$,

$$\psi_\pm(\boldsymbol{q}^{\boldsymbol{k}},\boldsymbol{p}^{\boldsymbol{k},\boldsymbol{\ell}}) = \sum_{\boldsymbol{m}} e^{\frac{\imath}{\varepsilon}\left(-\boldsymbol{p}^{\boldsymbol{k},\boldsymbol{\ell}}\cdot(\boldsymbol{y}^{\boldsymbol{m}}-\boldsymbol{q}^{\boldsymbol{k}})+\frac{\imath}{2}|\boldsymbol{y}^{\boldsymbol{m}}-\boldsymbol{q}^{\boldsymbol{k}}|^2\right)}$$

$$\times u_{\pm,0}(\boldsymbol{y}^{\boldsymbol{m}})r_\theta(|\boldsymbol{y}^{\boldsymbol{m}}-\boldsymbol{q}^{\boldsymbol{k}}|)\delta y_1\cdots\delta y_d, \quad (2.12)$$

where r_θ is a cutoff function such that $r_\theta = 1$ in the ball of radius $\theta > 0$ centered at the origin and $r_\theta = 0$ outside the ball.

Step 3. Solve (2.4)-(2.5) with the initial conditions

$$\boldsymbol{Q}_\pm(0,\boldsymbol{q^k},\boldsymbol{p^{k,\ell}}) = \boldsymbol{q^k}, \qquad \boldsymbol{P}_\pm(0,\boldsymbol{q^k},\boldsymbol{p^{k,\ell}}) = \boldsymbol{p^{k,\ell}},$$
$$a_\pm(0,\boldsymbol{q^k},\boldsymbol{p^{k,\ell}}) = 2^{d/2},$$

by a standard numerical integrator for ODE, for example the fourth-order Runge-Kutta scheme. Denote the numerical solutions as $(\boldsymbol{Q_\pm^{k,\ell}}, \boldsymbol{P_\pm^{k,\ell}})$ and $a_\pm^{\boldsymbol{k,\ell}}$.

Step 4. Reconstruct the solution by (2.10),

$$u^{\text{FGA}}(t,\boldsymbol{x^n}) = \sum_{\boldsymbol{k},\boldsymbol{\ell}} \left(\frac{a_+^{\boldsymbol{k,\ell}} r_\theta^+}{(2\pi\varepsilon)^{3d/2}} \psi_+(\boldsymbol{q^k},\boldsymbol{p^{k,\ell}}) e^{\frac{i}{\varepsilon}\boldsymbol{P_+^{k,\ell}}\cdot(\boldsymbol{x^n}-\boldsymbol{Q_+^{k,\ell}}) - \frac{1}{2\varepsilon}|\boldsymbol{x^n}-\boldsymbol{Q_+^{k,\ell}}|^2} \right.$$
$$\left. + \frac{a_-^{\boldsymbol{k,\ell}} r_\theta^-}{(2\pi\varepsilon)^{3d/2}} \psi_-(\boldsymbol{q^k},\boldsymbol{p^{k,\ell}}) e^{\frac{i}{\varepsilon}\boldsymbol{P_-^{k,\ell}}\cdot(\boldsymbol{x^n}-\boldsymbol{Q_-^{k,\ell}}) - \frac{1}{2\varepsilon}|\boldsymbol{x^n}-\boldsymbol{Q_-^{k,\ell}}|^2} \right)$$
$$\times \delta q_1 \cdots \delta q_d \delta p_1 \cdots \delta p_d,$$

(2.13)

where $r_\theta^\pm = r_\theta(|\boldsymbol{x^n} - \boldsymbol{Q_\pm^{k,\ell}}|)$.

REMARK 2.2.

1. In setting up the meshes, we assume that the initial condition (1.2) either has compact support or decays sufficiently fast to zero as $\boldsymbol{x} \to \infty$ so that we only need a finite number of mesh points in physical space.

2. The role of the truncation function r_θ is to save computational cost, since although a Gaussian function is not localized, it decays quickly away from the center. In practice we take $\theta = \mathcal{O}(\sqrt{\varepsilon})$, the same order as the width of each Gaussian, when we evaluate (2.12) and (2.13) numerically.

3. There are two types of errors present in the method. The first type comes from the asymptotic approximation to the wave equation. This error can *not* be reduced unless one includes higher order corrections. The other type is the numerical error which comes from two sources: one is from the ODE numerical integrator; the other is from the discrete approximations of integrals (2.10) and (2.11). It *can* be reduced by either taking small mesh size and time step or using higher order numerical methods.

4. Note that the assumption that the initial conditions are either compactly supported or decay quickly implies that the values on the boundary are zero (or close to zero). Then (2.12) and (2.13) are the trapezoidal rules to approximate (2.11) and (2.10). Notice that, due to the Gaussian factor, the integrand functions in (2.11) and (2.10) are exponentially small unless $\boldsymbol{x} - \boldsymbol{Q}$ and $\boldsymbol{y} - \boldsymbol{q}$ are on the order of $\mathcal{O}(\varepsilon^{1/2})$, which implies their derivatives with respect to $\boldsymbol{y}$, $\boldsymbol{q}$, $\boldsymbol{p}$ are of the order $\mathcal{O}(\varepsilon^{-1/2})$. This suggests $\boldsymbol{\delta y}$, $\boldsymbol{\delta q}$, $\boldsymbol{\delta p}$ should be taken as the size of $\mathcal{O}(\sqrt{\varepsilon})$. Hence N_y and N_q are of order $\mathcal{O}(\varepsilon^{-d/2})$. As illustrated in Figure 2.1, N_p is usually taken as $\mathcal{O}\left(\frac{\sqrt{\varepsilon}}{\min_j \delta p_j}\right)$, which is of order $\mathcal{O}(1)$. N_x is not constrained by ε, and is only determined by how well represented one wants the final solution.

5. Step 2 and 4 can be expedited by making use of the discrete fast Gaussian transform, as in [23, 22].

3. Asymptotic derivation

We now derive the formulation shown in Section 2 using asymptotic analysis. We start with the following ansatz for the wave Equation (1.1):

$$
\begin{aligned}
u(t,\boldsymbol{x}) = &\frac{1}{(2\pi\varepsilon)^{3d/2}} \int_{\mathbb{R}^{3d}} a_+(t,\boldsymbol{q},\boldsymbol{p}) e^{\frac{\imath}{\varepsilon}\Phi_+(t,\boldsymbol{x},\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})} u_{+,0}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q} \\
&+ \frac{1}{(2\pi\varepsilon)^{3d/2}} \int_{\mathbb{R}^{3d}} a_-(t,\boldsymbol{q},\boldsymbol{p}) e^{\frac{\imath}{\varepsilon}\Phi_-(t,\boldsymbol{x},\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})} u_{-,0}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q},
\end{aligned}
\tag{3.1}
$$

where $\Phi_\pm$ are given by

$$
\begin{aligned}
\Phi_\pm(t,\boldsymbol{x},\boldsymbol{y},\boldsymbol{q},\boldsymbol{p}) = S_\pm(t,\boldsymbol{q},\boldsymbol{p}) + \boldsymbol{P}_\pm(t,\boldsymbol{q},\boldsymbol{p})\cdot(\boldsymbol{x}-\boldsymbol{Q}_\pm(t,\boldsymbol{q},\boldsymbol{p})) - \boldsymbol{p}\cdot(\boldsymbol{y}-\boldsymbol{q}) \\
+\frac{\imath}{2}|\boldsymbol{x}-\boldsymbol{Q}_\pm(t,\boldsymbol{q},\boldsymbol{p})|^2 + \frac{\imath}{2}|\boldsymbol{y}-\boldsymbol{q}|^2. \quad (3.2)
\end{aligned}
$$

The initial conditions are taken as

$$
\begin{aligned}
\boldsymbol{Q}_\pm(0,\boldsymbol{q},\boldsymbol{p}) &= \boldsymbol{q}, & \boldsymbol{P}_\pm(0,\boldsymbol{q},\boldsymbol{p}) &= \boldsymbol{p}, \\
S_\pm(0,\boldsymbol{q},\boldsymbol{p}) &= 0, & a_\pm(0,\boldsymbol{q},\boldsymbol{p}) &= 2^{d/2}.
\end{aligned}
\tag{3.3}
$$

The subscript $\pm$ indicates the two branches that correspond to two different Hamiltonians,

$$
H_+(\boldsymbol{Q}_+,\boldsymbol{P}_+) = c(\boldsymbol{Q}_+)|\boldsymbol{P}_+|, \qquad H_-(\boldsymbol{Q}_-,\boldsymbol{P}_-) = -c(\boldsymbol{Q}_-)|\boldsymbol{P}_-|.
\tag{3.4}
$$

$\boldsymbol{P}_\pm$ and $\boldsymbol{Q}_\pm$ satisfy the equation of motion given by the Hamiltonian $H_\pm$:

$$
\begin{cases}
\partial_t \boldsymbol{Q}_\pm = \partial_{\boldsymbol{P}_\pm} H_\pm = \pm c\dfrac{\boldsymbol{P}_\pm}{|\boldsymbol{P}_\pm|}, \\
\partial_t \boldsymbol{P}_\pm = -\partial_{\boldsymbol{Q}_\pm} H_\pm = \mp\partial_{\boldsymbol{Q}_\pm} c|\boldsymbol{P}_\pm|.
\end{cases}
\tag{3.5}
$$

By plugging (3.1) into (1.1), the leading order terms show that the evolution of $S_\pm$ simply satisfies

$$
\partial_t S_\pm = 0.
\tag{3.6}
$$

This implies $S_\pm(t,\boldsymbol{q},\boldsymbol{p}) = 0$. Hence we omit the terms $S_\pm$ in Section 2 and later calculations.

Before proceeding further, let us state some lemmas that will be used.

LEMMA 3.1. *For $u \in L^2(\mathbb{R}^d)$, it holds*

$$
u(\boldsymbol{x}) = \frac{1}{(2\pi\varepsilon)^{3d/2}} \int_{\mathbb{R}^{3d}} 2^{d/2} e^{\frac{\imath}{\varepsilon}\Phi_\pm(0,\boldsymbol{x},\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})} u(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q}.
\tag{3.7}
$$

Proof. By the initial conditions (3.3),

$$
\Phi_\pm(0,\boldsymbol{x},\boldsymbol{y},\boldsymbol{q},\boldsymbol{p}) = \boldsymbol{p}\cdot(\boldsymbol{x}-\boldsymbol{q}) - \boldsymbol{p}\cdot(\boldsymbol{y}-\boldsymbol{q}) + \frac{\imath}{2}|\boldsymbol{x}-\boldsymbol{q}|^2 + \frac{\imath}{2}|\boldsymbol{y}-\boldsymbol{q}|^2.
\tag{3.8}
$$

Therefore, (3.7) is just the standard wave packet decomposition in disguise (see for example [4]). □

The proof of the following important lemma follows the one of Lemma 3 in [27].

LEMMA 3.2. *For any vector $\boldsymbol{a}(t,\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})$ and matrix $M(t,\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})$ in the Schwartz class, viewed as functions of $(\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})$, we have*

$$\boldsymbol{a}(t,\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})\cdot(\boldsymbol{x}-\boldsymbol{Q})\sim-\varepsilon\partial_{z_k}(a_j Z_{jk}^{-1}),\tag{3.9}$$

and

$$(\boldsymbol{x}-\boldsymbol{Q})\cdot M(t,\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})(\boldsymbol{x}-\boldsymbol{Q})\sim\varepsilon\partial_{z_l}Q_j M_{jk}Z_{kl}^{-1}+\varepsilon^2\partial_{z_m}\left(\partial_{z_l}(M_{jk}Z_{kl}^{-1})Z_{jm}^{-1}\right),\tag{3.10}$$

where Einstein's summation convention has been used.
Moreover, for any multi-index α such that $|\alpha|\geq 3$,

$$(\boldsymbol{x}-\boldsymbol{Q})^\alpha\sim\mathcal{O}(\varepsilon^{|\alpha|-1}).\tag{3.11}$$

Here we use the notation $f\sim g$ to mean that

$$\int_{\mathbb{R}^{3d}}f e^{\frac{\imath}{\varepsilon}\Phi_\pm}\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q}=\int_{\mathbb{R}^{3d}}g e^{\frac{\imath}{\varepsilon}\Phi_\pm}\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q}.\tag{3.12}$$

Proof.   Since the proof is exactly the same for the cases of Φ_+ and Φ_-, we omit the subscript $\pm$ for simplicity. As $\boldsymbol{a}$ and M are in the Schwartz class, all the manipulations below are justified.

Observe that at $t=0$,

$$-(\partial_{\boldsymbol{q}}\boldsymbol{Q})\boldsymbol{P}+\boldsymbol{p}=0,\qquad(\partial_{\boldsymbol{p}}\boldsymbol{Q})\boldsymbol{P}=0.$$

Using (3.5), we have

$$\begin{aligned}\partial_t\left(-(\partial_{\boldsymbol{q}}\boldsymbol{Q})\boldsymbol{P}+\boldsymbol{p}\right)&=-\partial_{\boldsymbol{q}}\left(\partial_t\boldsymbol{Q}\right)\boldsymbol{P}-\partial_{\boldsymbol{q}}\boldsymbol{Q}\partial_t\boldsymbol{P}\\&=-\partial_{\boldsymbol{q}}\left(c\frac{\boldsymbol{P}}{|\boldsymbol{P}|}\right)\boldsymbol{P}+\partial_{\boldsymbol{q}}\boldsymbol{Q}\partial_{\boldsymbol{Q}}c|\boldsymbol{P}|\\&=0.\end{aligned}$$

Analogously we have $\partial_t\left((\partial_{\boldsymbol{p}}\boldsymbol{Q})\boldsymbol{P}\right)=0$. Therefore for all $t>0$,

$$-(\partial_{\boldsymbol{q}}\boldsymbol{Q})\boldsymbol{P}+\boldsymbol{p}=0,\qquad(\partial_{\boldsymbol{p}}\boldsymbol{Q})\boldsymbol{P}=0.$$

Then straightforward calculations yield

$$\begin{aligned}\partial_{\boldsymbol{q}}\Phi&=(\partial_{\boldsymbol{q}}\boldsymbol{P}-\imath\partial_{\boldsymbol{q}}\boldsymbol{Q})(\boldsymbol{x}-\boldsymbol{Q})-\imath(\boldsymbol{y}-\boldsymbol{q}),\\\partial_{\boldsymbol{p}}\Phi&=(\partial_{\boldsymbol{p}}\boldsymbol{P}-\imath\partial_{\boldsymbol{p}}\boldsymbol{Q})(\boldsymbol{x}-\boldsymbol{Q})-(\boldsymbol{y}-\boldsymbol{q}),\end{aligned}$$

which implies that

$$\imath\partial_{\boldsymbol{z}}\Phi=Z(\boldsymbol{x}-\boldsymbol{Q}),\tag{3.13}$$

where $\partial_{\boldsymbol{z}}$ and Z are defined in (2.6). Note that Z can be rewritten as

$$Z=\partial_{\boldsymbol{z}}(\boldsymbol{Q}+\imath\boldsymbol{P})=\begin{pmatrix}\imath I & I\end{pmatrix}\begin{pmatrix}\partial_{\boldsymbol{q}}\boldsymbol{Q} & \partial_{\boldsymbol{q}}\boldsymbol{P}\\\partial_{\boldsymbol{p}}\boldsymbol{Q} & \partial_{\boldsymbol{p}}\boldsymbol{P}\end{pmatrix}\begin{pmatrix}-\imath I\\I\end{pmatrix},$$

where I stands for the $d\times d$ identity matrix. Therefore, define

$$F=\begin{pmatrix}\partial_{\boldsymbol{q}}\boldsymbol{Q} & \partial_{\boldsymbol{q}}\boldsymbol{P}\\\partial_{\boldsymbol{p}}\boldsymbol{Q} & \partial_{\boldsymbol{p}}\boldsymbol{P}\end{pmatrix}.$$

Then

$$ZZ^* = (\imath I\ I)\, F \begin{pmatrix} I & -\imath I \\ \imath I & I \end{pmatrix} F^{\mathrm{T}} \begin{pmatrix} -\imath I \\ I \end{pmatrix}$$

$$= (\imath I\ I)\, F F^{\mathrm{T}} \begin{pmatrix} -\imath I \\ I \end{pmatrix} + (\imath I\ I)\, F \begin{pmatrix} 0 & -\imath I \\ \imath I & 0 \end{pmatrix} F^{\mathrm{T}} \begin{pmatrix} -\imath I \\ I \end{pmatrix}$$

$$= (\imath I\ I)\, F F^{\mathrm{T}} \begin{pmatrix} -\imath I \\ I \end{pmatrix} + 2I.$$

In the last equality, we have used the fact that

$$F \begin{pmatrix} 0 & -\imath I \\ \imath I & 0 \end{pmatrix} F^{\mathrm{T}} = \begin{pmatrix} 0 & -\imath I \\ \imath I & 0 \end{pmatrix},$$

as F is the Jacobian of the Hamiltonian flow mapping $(\boldsymbol{p}, \boldsymbol{q}) \to (\boldsymbol{P}, \boldsymbol{Q})$. Moreover, F is invertible, and hence $F F^{\mathrm{T}}$ is positive definite. Therefore ZZ^* is positive definite for all t, which implies Z is invertible and

$$(\boldsymbol{x} - \boldsymbol{Q}) = \imath Z^{-1} \partial_{\boldsymbol{z}} \Phi. \tag{3.14}$$

Using (3.14), one has

$$\int_{\mathbb{R}^{3d}} \boldsymbol{a} \cdot (\boldsymbol{x} - \boldsymbol{Q}) e^{\frac{\imath}{\varepsilon}\Phi} \, \mathrm{d}\boldsymbol{y}\, \mathrm{d}\boldsymbol{p}\, \mathrm{d}\boldsymbol{q} = \varepsilon \int_{\mathbb{R}^{3d}} a_j Z_{jk}^{-1} \left(\frac{\imath}{\varepsilon} \partial_{z_k} \Phi \right) e^{\frac{\imath}{\varepsilon}\Phi} \, \mathrm{d}\boldsymbol{y}\, \mathrm{d}\boldsymbol{p}\, \mathrm{d}\boldsymbol{q}$$

$$= -\varepsilon \int_{\mathbb{R}^{3d}} \partial_{z_k} \left(a_j Z_{jk}^{-1} \right) e^{\frac{\imath}{\varepsilon}\Phi} \, \mathrm{d}\boldsymbol{y}\, \mathrm{d}\boldsymbol{p}\, \mathrm{d}\boldsymbol{q},$$

where the last equality is obtained from integration by parts. This proves (3.9).

Making use of (3.9) twice produces (3.10):

$$(\boldsymbol{x} - \boldsymbol{Q}) \cdot M (\boldsymbol{x} - \boldsymbol{Q}) = (x - Q)_j M_{jk} (x - Q)_k$$

$$\sim -\varepsilon \partial_{z_l} \left((x - Q)_j M_{jk} Z_{kl}^{-1} \right)$$

$$= \varepsilon \partial_{z_l} Q_j M_{jk} Z_{kl}^{-1} - \varepsilon (x - Q)_j \partial_{z_l} (M_{jk} Z_{kl}^{-1})$$

$$\sim \varepsilon \partial_{z_l} Q_j M_{jk} Z_{kl}^{-1} + \varepsilon^2 \partial_{z_m} \left(\partial_{z_l} (M_{jk} Z_{kl}^{-1}) Z_{jm}^{-1} \right).$$

By induction it is easy to see that (3.11) is true.   □

3.1. Initial value decomposition.   By (3.2) and (3.5) we obtain that

$$\partial_t \Phi_\pm = -\boldsymbol{P}_\pm \cdot \partial_t \boldsymbol{Q}_\pm + (\partial_t \boldsymbol{P}_\pm - \imath \partial_t \boldsymbol{Q}_\pm) \cdot (\boldsymbol{x} - \boldsymbol{Q}_\pm)$$

$$= \mp c |\boldsymbol{P}_\pm| \mp (\boldsymbol{x} - \boldsymbol{Q}_\pm) \cdot \left(|\boldsymbol{P}_\pm| \partial_{\boldsymbol{Q}_\pm} c + \imath \frac{\boldsymbol{P}_\pm}{|\boldsymbol{P}_\pm|} c \right), \tag{3.15}$$

and in particular for $t = 0$,

$$\partial_t \Phi_\pm (0, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{q}, \boldsymbol{p}) = \mp c |\boldsymbol{p}| \mp (\boldsymbol{x} - \boldsymbol{q}) \cdot \left(|\boldsymbol{p}| \partial_{\boldsymbol{q}} c + \imath \frac{\boldsymbol{p}}{|\boldsymbol{p}|} c \right). \tag{3.16}$$

The ansatz (3.1) shows that

$$u(0, \boldsymbol{x}) = \frac{1}{(2\pi\varepsilon)^{3d/2}} \int_{\mathbb{R}^{3d}} a_+(0, \boldsymbol{q}, \boldsymbol{p}) e^{\frac{\imath}{\varepsilon}\Phi_+(0, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{q}, \boldsymbol{p})} u_{+,0}(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y}\, \mathrm{d}\boldsymbol{p}\, \mathrm{d}\boldsymbol{q}$$

$$+ \frac{1}{(2\pi\varepsilon)^{3d/2}} \int_{\mathbb{R}^{3d}} a_-(0, \boldsymbol{p}, \boldsymbol{q}) e^{\frac{\imath}{\varepsilon}\Phi_-(0, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{q}, \boldsymbol{p})} u_{-,0}(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y}\, \mathrm{d}\boldsymbol{p}\, \mathrm{d}\boldsymbol{q}, \tag{3.17}$$

and

$$\partial_t u(0,\boldsymbol{x}) = \frac{1}{(2\pi\varepsilon)^{3d/2}} \int_{\mathbb{R}^{3d}} \left(\partial_t a_+ + \frac{\imath a_+}{\varepsilon}\partial_t\Phi_+\right)e^{\frac{\imath}{\varepsilon}\Phi_+(0,\boldsymbol{x},\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})}u_{+,0}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q}$$
$$+ \frac{1}{(2\pi\varepsilon)^{3d/2}} \int_{\mathbb{R}^{3d}} \left(\partial_t a_- + \frac{\imath a_-}{\varepsilon}\partial_t\Phi_-\right)e^{\frac{\imath}{\varepsilon}\Phi_-(0,\boldsymbol{x},\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})}u_{-,0}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q}. \tag{3.18}$$

We take

$$a_\pm(0,\boldsymbol{q},\boldsymbol{p}) = 2^{d/2}, \tag{3.19}$$
$$u_{+,0}(\boldsymbol{x}) = A_+(\boldsymbol{x})e^{\frac{\imath}{\varepsilon}S_0(\boldsymbol{x})}, \tag{3.20}$$
$$u_{-,0}(\boldsymbol{x}) = A_-(\boldsymbol{x})e^{\frac{\imath}{\varepsilon}S_0(\boldsymbol{x})}, \tag{3.21}$$

with

$$A_\pm(\boldsymbol{x}) = \frac{1}{2}\left(A_0(\boldsymbol{x}) \pm \frac{\imath B_0(\boldsymbol{x})}{c(\boldsymbol{x})|\partial_{\boldsymbol{x}} S_0(\boldsymbol{x})|}\right). \tag{3.22}$$

We next show that this will approximate the initial condition to the leading order in ε.

Substituting (3.19)-(3.22) into (3.17) and using Lemma 3.1, we easily confirm that

$$u(0,\boldsymbol{x}) = u_{+,0}(\boldsymbol{x}) + u_{-,0}(\boldsymbol{x}) = A_0(\boldsymbol{x})e^{\frac{\imath}{\varepsilon}S_0(\boldsymbol{x})}.$$

For the initial velocity, we substitute (3.19)-(3.22) into (3.18) and keep only the leading order terms in ε. According to Lemma 3.2, only the term $\mp c|\boldsymbol{p}|$ in $\partial_t\Phi_\pm$ will contribute to the leading order, since the other terms that contain $(\boldsymbol{x}-\boldsymbol{q})$ are $\mathcal{O}(\varepsilon)$. Hence,

$$\partial_t u(0,\boldsymbol{x}) = -\frac{2^{d/2}}{(2\pi\varepsilon)^{3d/2}} \int_{\mathbb{R}^{3d}} \frac{\imath}{\varepsilon}c(\boldsymbol{q})|\boldsymbol{p}|e^{\frac{\imath}{\varepsilon}\Phi_+(0,\boldsymbol{x},\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})}u_{+,0}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q}$$
$$+ \frac{2^{d/2}}{(2\pi\varepsilon)^{3d/2}} \int_{\mathbb{R}^{3d}} \frac{\imath}{\varepsilon}c(\boldsymbol{q})|\boldsymbol{p}|e^{\frac{\imath}{\varepsilon}\Phi_-(0,\boldsymbol{x},\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})}u_{-,0}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q} + \mathcal{O}(1).$$

Consider the integral

$$\int_{\mathbb{R}^d} c(\boldsymbol{q})|\boldsymbol{p}|e^{-\frac{\imath}{\varepsilon}\boldsymbol{p}\cdot(\boldsymbol{y}-\boldsymbol{q})-\frac{1}{2\varepsilon}|\boldsymbol{y}-\boldsymbol{q}|^2}A_\pm(\boldsymbol{y})e^{\frac{\imath}{\varepsilon}S_0(\boldsymbol{y})}\,\mathrm{d}\boldsymbol{y} = \int_{\mathbb{R}^d} c(\boldsymbol{q})|\boldsymbol{p}|A_\pm(\boldsymbol{y})e^{\frac{\imath}{\varepsilon}\Theta(\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})}\,\mathrm{d}\boldsymbol{y}.$$

The phase function Θ is given by

$$\Theta(\boldsymbol{y},\boldsymbol{q},\boldsymbol{p}) = -\boldsymbol{p}\cdot(\boldsymbol{y}-\boldsymbol{q}) + \frac{\imath}{2}|\boldsymbol{y}-\boldsymbol{q}|^2 + S_0(\boldsymbol{y}).$$

Clearly, $\mathfrak{Im}\,\Theta \geq 0$ and $\mathfrak{Im}\,\Theta = 0$ if and only if $\boldsymbol{y} = \boldsymbol{q}$. The derivatives of Θ with respect to $\boldsymbol{y}$ are

$$\partial_{\boldsymbol{y}}\Theta = -\boldsymbol{p} + \partial_{\boldsymbol{y}}S_0(\boldsymbol{y}) + \imath(\boldsymbol{y}-\boldsymbol{q}),$$
$$\partial_{\boldsymbol{y}}^2\Theta = \partial_{\boldsymbol{y}}^2 S_0(\boldsymbol{y}) + iI.$$

Hence, the first derivative vanishes only when $\boldsymbol{y} = \boldsymbol{q}$ and $\boldsymbol{p} = \partial_{\boldsymbol{y}}S_0(\boldsymbol{y})$, and $\det\partial_{\boldsymbol{y}}^2\Theta \neq 0$. Therefore, we can apply stationary phase approximation with complex phase (see for

example [7]) to conclude, for $(\boldsymbol{q},\boldsymbol{p})\in\mathbb{R}^{2d}$,

$$\int_{\mathbb{R}^d} c(\boldsymbol{q})|\boldsymbol{p}|e^{-\frac{\imath}{\varepsilon}\boldsymbol{p}\cdot(\boldsymbol{y}-\boldsymbol{q})-\frac{1}{2\varepsilon}|\boldsymbol{y}-\boldsymbol{q}|^2}A_\pm(\boldsymbol{y})e^{\frac{\imath}{\varepsilon}S_0(\boldsymbol{y})}\,\mathrm{d}\boldsymbol{y}$$

$$=\int_{\mathbb{R}^d} c(\boldsymbol{y})|\partial_{\boldsymbol{y}}S_0(\boldsymbol{y})|e^{-\frac{\imath}{\varepsilon}\boldsymbol{p}\cdot(\boldsymbol{y}-\boldsymbol{q})-\frac{1}{2\varepsilon}|\boldsymbol{y}-\boldsymbol{q}|^2}A_\pm(\boldsymbol{y})e^{\frac{\imath}{\varepsilon}S_0(\boldsymbol{y})}\,\mathrm{d}\boldsymbol{y}+\mathcal{O}(\varepsilon).$$

Therefore,

$$\partial_t u(0,\boldsymbol{x})=-\frac{2^{d/2}}{(2\pi\varepsilon)^{3d/2}}\int_{\mathbb{R}^{3d}}\frac{\imath}{\varepsilon}c(\boldsymbol{y})|\partial_{\boldsymbol{y}}S_0(\boldsymbol{y})|e^{\frac{\imath}{\varepsilon}\Phi_+(0,\boldsymbol{x},\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})}u_{+,0}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q}$$

$$+\frac{2^{d/2}}{(2\pi\varepsilon)^{3d/2}}\int_{\mathbb{R}^{3d}}\frac{\imath}{\varepsilon}c(\boldsymbol{y})|\partial_{\boldsymbol{y}}S_0(\boldsymbol{y})|e^{\frac{\imath}{\varepsilon}\Phi_-(0,\boldsymbol{x},\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})}u_{-,0}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q}+\mathcal{O}(1). \quad (3.23)$$

Substitute (3.22) into (3.23) and use Lemma 3.1, then

$$\partial_t u(0,\boldsymbol{x})=\frac{1}{\varepsilon}B_0(\boldsymbol{x})e^{\frac{\imath}{\varepsilon}S_0(\boldsymbol{x})}+\mathcal{O}(1),$$

which agrees with (1.2) to the leading order.

3.2. Derivation of the evolution equation of $a_\pm$.   In order to derive the evolution equation for the weight function a, we carry out the asymptotic analysis of the wave Equation (1.1) using the ansatz (3.1) in this section. As the Equation (1.1) is linear, we can deal with the two branches separately. In the following, we only deal with the "+" branch that corresponds to H_+, and the other is completely analogous. For simplicity, we drop the subscript "+" in the notations.

Substituting (3.1) into the Equation (1.1) (keeping only the "+" branch) gives

$$\partial_t^2 u=\frac{1}{(2\pi\varepsilon)^{3d/2}}\int_{\mathbb{R}^{3d}}\Big(\partial_t^2 a+2\frac{\imath}{\varepsilon}\partial_t a\partial_t\Phi+\frac{\imath}{\varepsilon}a\partial_t^2\Phi-\frac{1}{\varepsilon^2}a(\partial_t\Phi)^2\Big)e^{\imath\Phi/\varepsilon}u_0\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q},$$

and

$$\Delta u=\frac{1}{(2\pi\varepsilon)^{3d/2}}\int_{\mathbb{R}^{3d}}\Big(\frac{\imath}{\varepsilon}\Delta\Phi-\frac{1}{\varepsilon^2}(\partial_{\boldsymbol{x}}\Phi\cdot\partial_{\boldsymbol{x}}\Phi)\Big)ae^{\imath\Phi/\varepsilon}u_0\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q}.$$

Squaring both sides of (3.15) yields

$$(\partial_t\Phi)^2=c^2|\boldsymbol{P}|^2+\Big((\boldsymbol{x}-\boldsymbol{Q})\cdot\Big(|\boldsymbol{P}|\partial_{\boldsymbol{Q}}c+\imath c\frac{\boldsymbol{P}}{|\boldsymbol{P}|}\Big)\Big)^2$$

$$+2c|\boldsymbol{P}|(\boldsymbol{x}-\boldsymbol{Q})\cdot\Big(|\boldsymbol{P}|\partial_{\boldsymbol{Q}}c+\imath c\frac{\boldsymbol{P}}{|\boldsymbol{P}|}\Big). \quad (3.24)$$

Differentiating (3.15) with respect to t, one has

$$\partial_t^2\Phi=-\partial_t(|\boldsymbol{P}|c)+\partial_t\boldsymbol{Q}\cdot\Big(|\boldsymbol{P}|\partial_{\boldsymbol{Q}}c+\imath\frac{\boldsymbol{P}}{|\boldsymbol{P}|}c\Big)$$

$$-(\boldsymbol{x}-\boldsymbol{Q})\cdot\Big(\partial_{\boldsymbol{Q}}c\frac{\boldsymbol{P}\cdot\partial_t\boldsymbol{P}}{|\boldsymbol{P}|}+\partial_{\boldsymbol{Q}}^2c\cdot\partial_t\boldsymbol{Q}|\boldsymbol{P}| \qquad\qquad (3.25)$$

$$+\imath c\frac{\partial_t\boldsymbol{P}}{|\boldsymbol{P}|}-\imath c\boldsymbol{P}\frac{\boldsymbol{P}\cdot\partial_t\boldsymbol{P}}{|\boldsymbol{P}|^3}+\imath\frac{\boldsymbol{P}}{|\boldsymbol{P}|}\partial_{\boldsymbol{Q}}c\cdot\partial_t\boldsymbol{Q}\Big).$$

We simplify the last equation using (2.4),

$$\partial_t^2 \Phi = c\boldsymbol{P} \cdot \partial_{\boldsymbol{Q}} c + \imath c^2$$
$$- (\boldsymbol{x} - \boldsymbol{Q}) \cdot \left(-\partial_{\boldsymbol{Q}} c \boldsymbol{P} \cdot \partial_{\boldsymbol{Q}} c + c \partial_{\boldsymbol{Q}}^2 c \cdot \boldsymbol{P} - \imath c \partial_{\boldsymbol{Q}} c + 2\imath c \boldsymbol{P} \frac{\boldsymbol{P} \cdot \partial_{\boldsymbol{Q}} c}{|\boldsymbol{P}|^2} \right). \tag{3.26}$$

Taking derivatives with respect to $\boldsymbol{x}$ produces

$$\partial_{\boldsymbol{x}} \Phi = \boldsymbol{P} + \imath (\boldsymbol{x} - \boldsymbol{Q}), \tag{3.27}$$

$$\partial_{\boldsymbol{x}} \Phi \cdot \partial_{\boldsymbol{x}} \Phi = |\boldsymbol{P}|^2 + 2\imath \boldsymbol{P} \cdot (\boldsymbol{x} - \boldsymbol{Q}) - |\boldsymbol{x} - \boldsymbol{Q}|^2, \tag{3.28}$$

and

$$\Delta \Phi = d\imath. \tag{3.29}$$

We next expand $c(\boldsymbol{x})$ around the point $\boldsymbol{Q}$,

$$c(\boldsymbol{x}) = c + \partial_{\boldsymbol{Q}} c \cdot (\boldsymbol{x} - \boldsymbol{Q}) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{Q}) \cdot \partial_{\boldsymbol{Q}}^2 c (\boldsymbol{x} - \boldsymbol{Q}) + \mathcal{O}(|\boldsymbol{x} - \boldsymbol{Q}|)^3, \tag{3.30}$$

and

$$c^2(\boldsymbol{x}) = c^2 + 2c\partial_{\boldsymbol{Q}} c \cdot (\boldsymbol{x} - \boldsymbol{Q}) + (\partial_{\boldsymbol{Q}} c \cdot (\boldsymbol{x} - \boldsymbol{Q}))^2$$
$$+ c(\boldsymbol{x} - \boldsymbol{Q}) \cdot \partial_{\boldsymbol{Q}}^2 c (\boldsymbol{x} - \boldsymbol{Q}) + \mathcal{O}(|\boldsymbol{x} - \boldsymbol{Q}|)^3. \tag{3.31}$$

The terms c, $\partial_{\boldsymbol{Q}} c$ and $\partial_{\boldsymbol{Q}}^2 c$ on the right hand sides are all evaluated at $\boldsymbol{Q}$.

Substituting all the above into the wave Equation (1.1) and keeping only the leading order terms gives

$$2\frac{\imath}{\varepsilon} \partial_t a(-c|\boldsymbol{P}|)u + \frac{\imath}{\varepsilon} a(c\boldsymbol{P} \cdot \partial_{\boldsymbol{Q}} c + \imath c^2)u$$
$$- \frac{1}{\varepsilon^2} a\left(2c(\boldsymbol{x} - \boldsymbol{Q}) \cdot (|\boldsymbol{P}|^2 \partial_{\boldsymbol{Q}} c + \imath c\boldsymbol{P}) + ((\boldsymbol{x} - \boldsymbol{Q}) \cdot (|\boldsymbol{P}| \partial_{\boldsymbol{Q}} c + \imath c\boldsymbol{P}/|\boldsymbol{P}|))^2 \right)u$$
$$- c^2\left(-\frac{1}{\varepsilon} ad - \frac{2\imath}{\varepsilon^2} a\boldsymbol{P} \cdot (\boldsymbol{x} - \boldsymbol{Q}) + \frac{1}{\varepsilon^2} a|\boldsymbol{x} - \boldsymbol{Q}|^2 \right)u$$
$$+ \frac{2}{\varepsilon^2} ac\partial_{\boldsymbol{Q}} c \cdot (\boldsymbol{x} - \boldsymbol{Q})(|\boldsymbol{P}|^2 + 2\imath \boldsymbol{P} \cdot (\boldsymbol{x} - \boldsymbol{Q}))u$$
$$+ \frac{1}{\varepsilon^2} a|\boldsymbol{P}|^2\left((\partial_{\boldsymbol{Q}} c \cdot (\boldsymbol{x} - \boldsymbol{Q}))^2 u + c(\boldsymbol{x} - \boldsymbol{Q}) \cdot \partial_{\boldsymbol{Q}}^2 c (\boldsymbol{x} - \boldsymbol{Q}) \right)u \sim \mathcal{O}(1).$$

After reorganizing the terms, we get

$$2\frac{\imath}{\varepsilon} c|\boldsymbol{P}| \partial_t a u \sim \frac{\imath}{\varepsilon} a(c\boldsymbol{P} \cdot \partial_{\boldsymbol{Q}} c - (d-1)c^2 \imath)u - \frac{1}{\varepsilon^2} a(\boldsymbol{x} - \boldsymbol{Q}) \cdot M(\boldsymbol{x} - \boldsymbol{Q})u, \tag{3.32}$$

where

$$M = (|\boldsymbol{P}| \partial_{\boldsymbol{Q}} c - \imath c\boldsymbol{P}/|\boldsymbol{P}|) \otimes (|\boldsymbol{P}| \partial_{\boldsymbol{Q}} c - \imath c\boldsymbol{P}/|\boldsymbol{P}|) + c^2 I$$
$$- |\boldsymbol{P}|^2 \partial_{\boldsymbol{Q}} c \otimes \partial_{\boldsymbol{Q}} c - |\boldsymbol{P}|^2 c \partial_{\boldsymbol{Q}}^2 c. \tag{3.33}$$

Lemma 3.2 shows that

$$a(\boldsymbol{x} - \boldsymbol{Q}) \cdot M(\boldsymbol{x} - \boldsymbol{Q})u \sim \varepsilon a\,\mathrm{tr}(Z^{-1} \partial_{\boldsymbol{z}} \boldsymbol{Q} M)u + \mathcal{O}(\varepsilon^2). \tag{3.34}$$

Therefore, to the leading order, we obtain the evolution equation of a:

$$
\partial_t a = \frac{a}{2}\left(\frac{\boldsymbol{P}}{|\boldsymbol{P}|}\cdot\partial_{\boldsymbol{Q}}c - \frac{(d-1)\imath}{|\boldsymbol{P}|}c\right)
$$
$$
+ \frac{a}{2}\operatorname{tr}\left(Z^{-1}\partial_{\boldsymbol{z}}\boldsymbol{Q}\left(2\frac{\boldsymbol{P}}{|\boldsymbol{P}|}\otimes\partial_{\boldsymbol{Q}}c - \frac{\imath c}{|\boldsymbol{P}|}\left(\frac{\boldsymbol{P}\otimes\boldsymbol{P}}{|\boldsymbol{P}|^2}-I\right)-\imath|\boldsymbol{P}|\partial_{\boldsymbol{Q}}^2 c\right)\right). \quad (3.35)
$$

Notice that

$$
\frac{\mathrm{d}Z}{\mathrm{d}t} = \partial_{\boldsymbol{z}}\left(\frac{\mathrm{d}\boldsymbol{Q}}{\mathrm{d}t}+\imath\frac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t}\right) = \partial_{\boldsymbol{z}}\left(c\frac{\boldsymbol{P}}{|\boldsymbol{P}|}-\imath\partial_{\boldsymbol{Q}}c|\boldsymbol{P}|\right)
$$
$$
= \partial_{\boldsymbol{z}}\boldsymbol{Q}\frac{\partial_{\boldsymbol{Q}}c\otimes\boldsymbol{P}}{|\boldsymbol{P}|} + c\partial_{\boldsymbol{z}}\boldsymbol{P}\left(\frac{I}{|\boldsymbol{P}|}-\frac{\boldsymbol{P}\otimes\boldsymbol{P}}{|\boldsymbol{P}|^3}\right) - \imath\partial_{\boldsymbol{z}}\boldsymbol{Q}\partial_{\boldsymbol{Q}}^2 c|\boldsymbol{P}| - \imath\partial_{\boldsymbol{z}}\boldsymbol{P}\frac{\boldsymbol{P}\otimes\partial_{\boldsymbol{Q}}c}{|\boldsymbol{P}|},
$$

and

$$
-\frac{(d-1)\imath}{|\boldsymbol{P}|}c = \operatorname{tr}\left(Z^{-1}(\partial_{\boldsymbol{z}}\boldsymbol{Q}+\imath\partial_{\boldsymbol{z}}\boldsymbol{P})\frac{\imath c}{|\boldsymbol{P}|}\left(\frac{\boldsymbol{P}\otimes\boldsymbol{P}}{|\boldsymbol{P}|^2}-I\right)\right).
$$

By using the fact that (3.34) has a quadratic form, one has

$$
\operatorname{tr}\left(Z^{-1}\partial_{\boldsymbol{z}}\boldsymbol{Q}\frac{\boldsymbol{P}}{|\boldsymbol{P}|}\otimes\partial_{\boldsymbol{Q}}c\right) = \operatorname{tr}\left(Z^{-1}\partial_{\boldsymbol{z}}\boldsymbol{Q}\frac{\partial_{\boldsymbol{Q}}c}{|\boldsymbol{P}|}\otimes\boldsymbol{P}\right).
$$

Hence (3.35) can be reformulated as

$$
\frac{\mathrm{d}a}{\mathrm{d}t} = a\frac{\boldsymbol{P}}{|\boldsymbol{P}|}\cdot\partial_{\boldsymbol{Q}}c + \frac{a}{2}\operatorname{tr}\left(Z^{-1}\frac{\mathrm{d}Z}{\mathrm{d}t}\right).
$$

This completes the asymptotic derivation. We remark that in the case of the time dependent Schrödinger equation, the asymptotics have been made rigorous in [27, 25].

4. Numerical examples

In this section, we give both one and two dimensional numerical examples to justify the accuracy of the frozen Gaussian approximation (FGA). Without loss of generality, we only consider the wave propagation determined by the "+" branch of (2.1) which implies that $B_0(\boldsymbol{x}) = -\imath c(\boldsymbol{x})|\nabla_{\boldsymbol{x}}S_0(\boldsymbol{x})|A_0(\boldsymbol{x})$ in (1.2).

4.1. One dimension.   Using one-dimensional examples in this section, we compare FGA with the Gaussian beam method (GBM) in both the accuracy and the performance when beams spread in GBM. We denote the solution of GBM as u^{GBM}, and summarize its discrete numerical formulation (only the "+" branch) as follows for readers' convenience ([24, 29, 14]),

$$
u^{\mathrm{GBM}}(t,x) = \sum_{j=1}^{N_{y_0}}\left(\frac{1}{2\pi\epsilon}\right)^{\frac{1}{2}}r_\theta(|x-y_j|)A(t,y_j)
$$
$$
\times\exp\left(\frac{\imath}{\varepsilon}(S(t,y_j)+\xi(t,y_j)(x-y_j)+M(t,y_j)(x-y_j)^2/2)\right)\delta y_0,
$$

ε	$1/2^6$	$1/2^7$	$1/2^8$
$\|u-u^{\mathrm{FGA}}\|_{\ell^\infty}$	1.12×10^{-1}	6.18×10^{-2}	2.51×10^{-2}
$\|u-u^{\mathrm{FGA}}\|_{\ell^2}$	6.05×10^{-2}	2.96×10^{-2}	1.19×10^{-2}
$\|u-u^{\mathrm{GBM}}\|_{\ell^\infty}$	7.15×10^{-1}	5.08×10^{-1}	3.36×10^{-1}
$\|u-u^{\mathrm{GBM}}\|_{\ell^2}$	3.26×10^{-1}	2.28×10^{-1}	1.47×10^{-1}

TABLE 4.1. *Example 4.1, the ℓ^∞ and ℓ^2 errors for FGA and GBM.*

and y_j, ξ, S, M, A satisfy

$$\frac{\mathrm{d}y_j}{\mathrm{d}t} = c(y_j)\frac{\xi}{|\xi|}, \qquad y_j(0) = y_0^j,$$

$$\frac{\mathrm{d}\xi}{\mathrm{d}t} = -\partial_{y_j}c(y_j)|\xi|, \qquad \xi(0) = \partial_{y_j}S_0(y_j),$$

$$\frac{\mathrm{d}S}{\mathrm{d}t} = 0, \qquad S(0) = S_0(y_j),$$

$$\frac{\mathrm{d}M}{\mathrm{d}t} = -2\partial_{y_j}c(y_j)\frac{\xi}{|\xi|}M - \partial_{y_j}^2 c(y_j)|\xi|, \qquad M(0) = \partial_{y_j}^2 S_0(y_j) + \imath,$$

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \frac{1}{2}\partial_{y_j}c(y_j)\frac{\xi}{|\xi|}A, \qquad A(0) = A_0(y_j),$$

where r_θ is the cutoff function, y_0^j's are the equidistant mesh points, δy_0 is the mesh size and N_{y_0} is the total number of the beams initially centered at y_0^j.

EXAMPLE 4.1. *The wave speed is $c(x) = x^2$. The initial conditions are*

$$u_0 = \exp\left(-100(x-0.5)^2\right)\exp\left(\frac{\imath x}{\varepsilon}\right),$$

$$\partial_t u_0 = -\frac{\imath x^2}{\varepsilon}\exp\left(-100(x-0.5)^2\right)\exp\left(\frac{\imath x}{\varepsilon}\right).$$

The final time is $T = 0.5$. We plot the real part of the wave field obtained by FGA compared with the true solution in Figure 4.1 for $\varepsilon = 1/64, 1/128, 1/256$. The true solution is computed by the finite difference method using the mesh size of $\delta x = 1/2^{12}$ and the time step of $\delta t = 1/2^{18}$ for domain $[0,2]$. Table 4.1 shows the ℓ^∞ and ℓ^2 errors of both the FGA solution u^{FGA} and the GBM solution u^{GBM}. The convergence orders in ε of the ℓ^∞ and ℓ^2 norms are 1.08 and 1.17 separately for FGA, and 0.54 and 0.57 for GBM. We observe a better accuracy order of FGA than GBM.

We choose $\delta t = 1/2^{11}$ for solving the ODEs and $\delta x = 1/2^{12}$ to construct the final solution in both FGA and GBM. In FGA, we take $\delta q = \delta p = \delta y = 1/2^7$, $N_q = 128$, $N_p = 45$ for $\varepsilon = 1/128, 1/256$ and $\delta q = \delta p = \delta y = 1/2^5$, $N_q = 32$, $N_p = 33$ for $\varepsilon = 1/64$. In GBM, we take $\delta y_0 = 1/2^7$, $N_{y_0} = 128$ for $\varepsilon = 1/128, 1/256$ and $\delta y_0 = 1/2^5$, $N_{y_0} = 32$ for $\varepsilon = 1/64$.

We remark that in this example the mesh sizes of p and q have been taken very small and N_p large enough to make sure that the error of FGA mostly comes from asymptotic expansion, but not from initial value decomposition, numerical integration of ODEs and so on. Such a choice of fine mesh is not necessary for the accuracy of FGA, as one can see in Example 4.3.
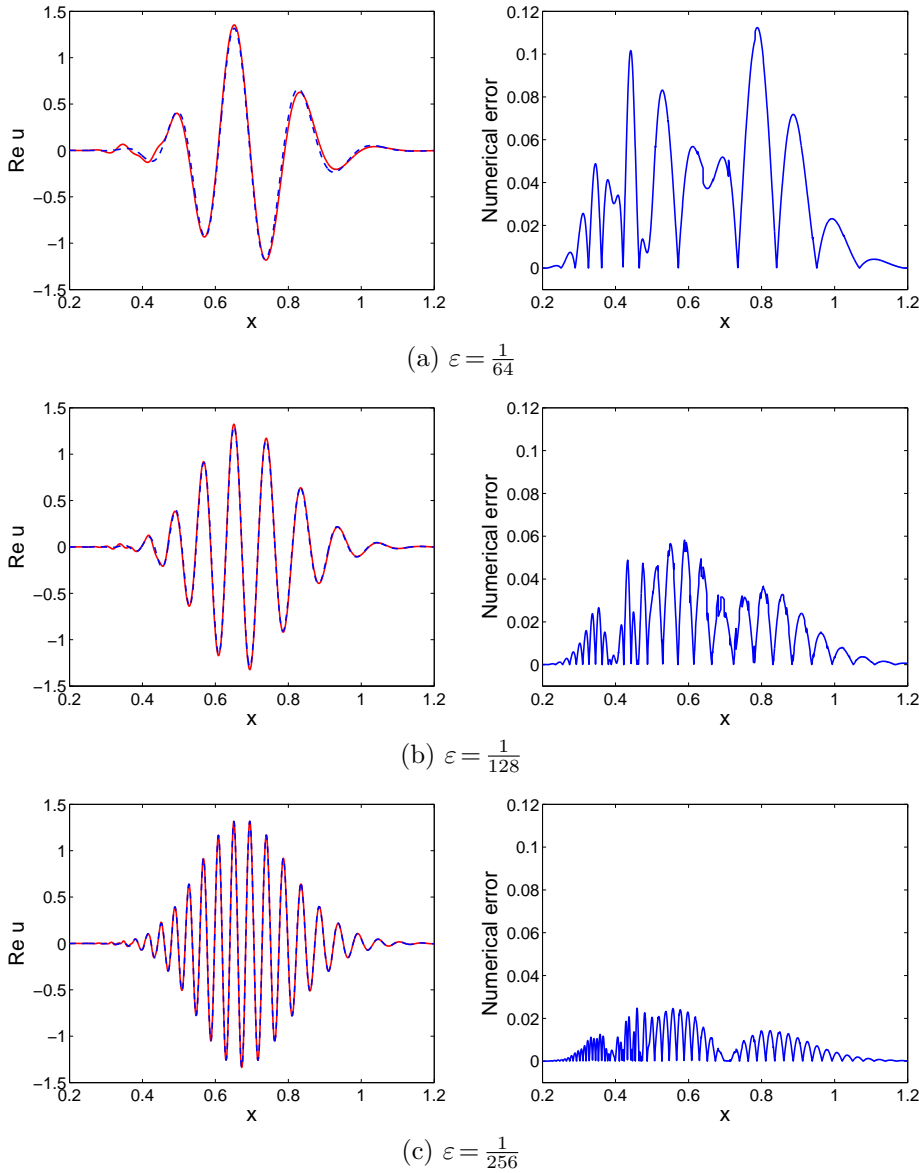
(a) $\varepsilon = \frac{1}{64}$

(b) $\varepsilon = \frac{1}{128}$

(c) $\varepsilon = \frac{1}{256}$

FIG. 4.1. *Example 4.1, the comparison of the true solution (solid line) and the solution by FGA (dashed line). Left: the real part of wave field; right: the errors between them.*

EXAMPLE 4.2. *The wave speed is $c(x) = x^2$. The initial conditions are*

$$u_0 = \exp\left(-\frac{(x-0.55)^2}{2\varepsilon}\right)\exp\left(\frac{\iota x}{\varepsilon}\right),$$

$$\partial_t u_0 = -\frac{\iota x^2}{\varepsilon}\exp\left(-\frac{(x-0.55)^2}{2\varepsilon}\right)\exp\left(\frac{\iota x}{\varepsilon}\right).$$

We use this example to illustrate the performances of FGA and GBM when the beams spread in GBM. The final time is $T = 1.0$ and $\varepsilon = 1/256$. Remark that the
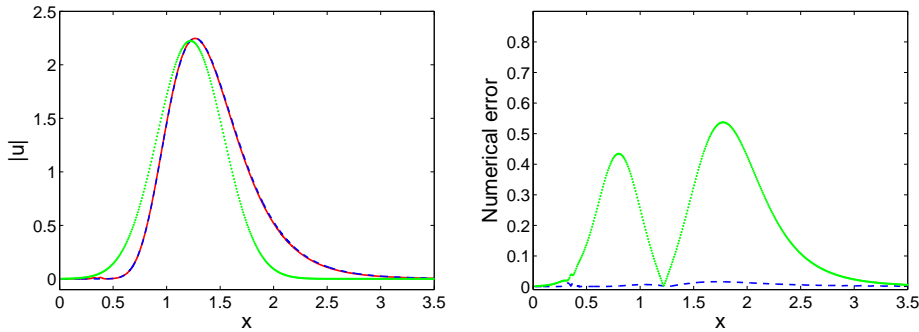
FIG. 4.2. *Example 4.2, the comparison of the true solution (solid line), the solution by FGA (dashed line) and the solution by GBM (dots) for $\varepsilon = \frac{1}{256}$. Left: the amplitude of wave field; right: the error between them (dashed line for FGA, dots for GBM).*
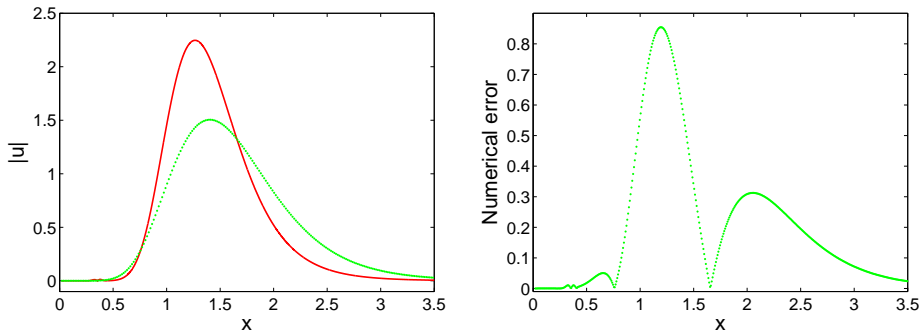

FIG. 4.3. *Example 4.2, the comparison of the true solution (solid line) and the solution by GBM using multiple Gaussian initial representation (dots) for $\varepsilon = \frac{1}{256}$. Left: the amplitude of wave field; right: the error between them.*
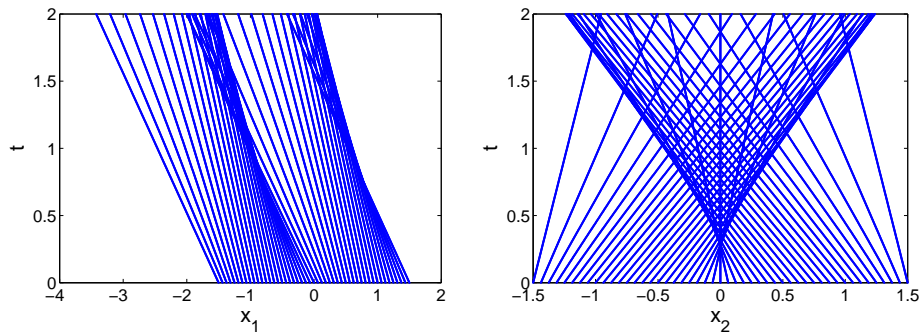
FIG. 4.4. *Example 4.3, a set of the characteristic lines develops the cusp caustic.*

initial condition is chosen as a single beam on purpose so that one can apply GBM without introducing any initial errors. The true solution is provided by the finite difference method using $\delta x = 1/2^{11}$ and $\delta t = 1/2^{17}$ for domain $[0,4]$. We take $\delta q = \delta p = \delta y = 1/2^7$, $N_q = 128$, $N_p = 45$ in FGA to make sure that the error in the initial value decomposition of FGA is very small. The time step is $\delta t = 1/2^{10}$ for solving the ODEs and the mesh size is chosen as $\delta x = 1/2^{11}$ to construct the final solution in both FGA
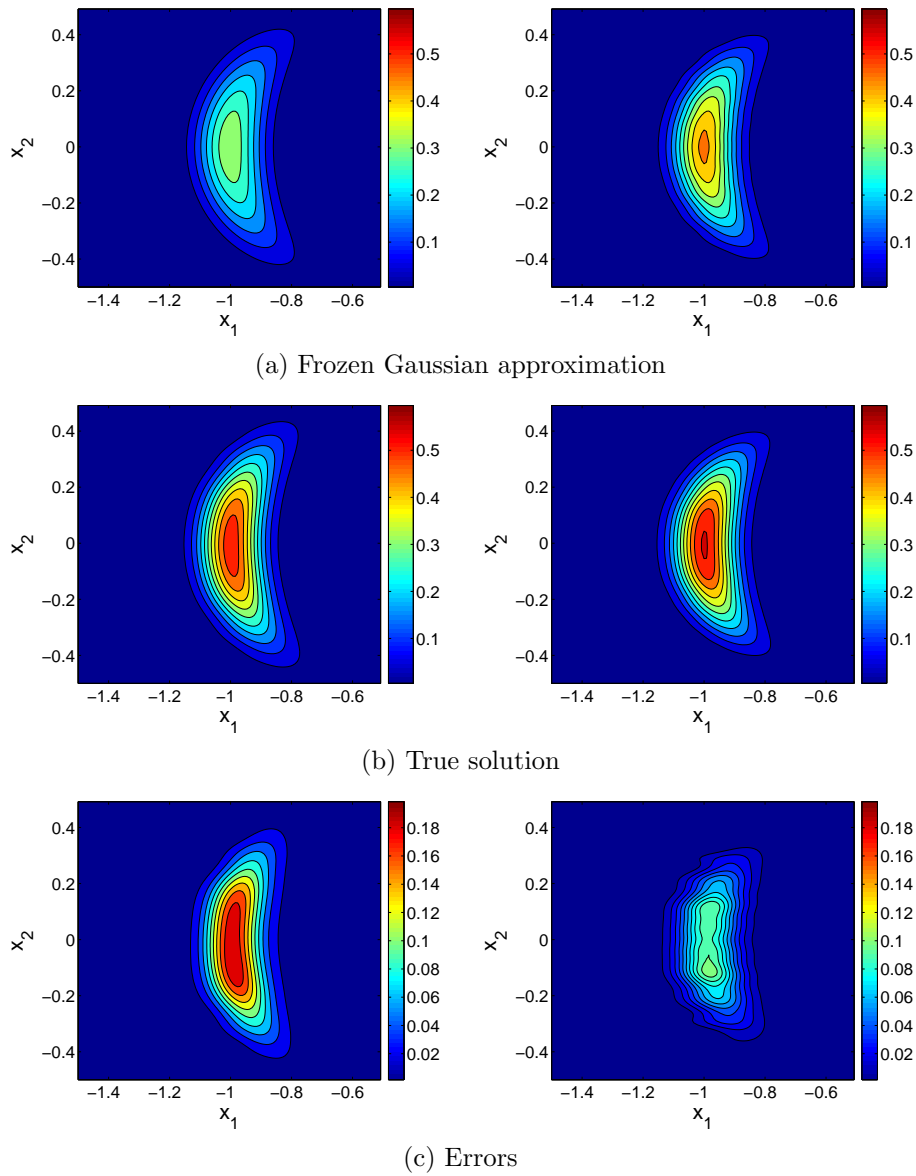
(a) Frozen Gaussian approximation

(b) True solution

(c) Errors

FIG. 4.5. *Example 4.3, the comparison of the true solution and the solution by FGA. Left: wave amplitude of $\varepsilon = \frac{1}{128}$; right: wave amplitude of $\varepsilon = \frac{1}{256}$.*

and GBM.

Figure 4.2 compares the amplitudes of the wave field computed by FGA and GBM, and the true solution. One can see that the beam has spread severely in GBM. The results confirm that FGA has a good performance even when the beam spreads, while GBM does not. Moreover, it does not help improving the accuracy if one uses more Gaussian beams to approximate the initial condition in GBM as shown in Figure 4.3, where $N_{y_0} = 128$ beams are used initially and $\delta y_0 = 1/2^7$. Remark that GBM can still give good approximation around beam center where Taylor expansion does not introduce large errors. This can be seen around $x = 1.2$ in Figure 4.2.

4.2. Two dimension.

EXAMPLE 4.3. *The wave speed is* $c(x_1, x_2) = 1$. *The initial conditions are*

$$u_0 = \exp\big(-100(x_1^2 + x_2^2)\big) \exp\Big(\frac{\imath}{\varepsilon}(-x_1 + \cos(2x_2))\Big),$$

$$\partial_t u_0 = -\frac{\imath}{\varepsilon}\sqrt{1 + 4\sin^2(2x_2)}\, \exp\big(-100(x_1^2 + x_2^2)\big)\exp\Big(\frac{\imath}{\varepsilon}(-x_1 + \cos(2x_2))\Big).$$

This example presents the cusp caustics shown in Figure 4.4. The final time is $T = 1.0$. The true solution is given by the spectral method using the mesh $\delta x_1 = \delta x_2 = 1/512$ for domain $[-1.5, 0.5] \times [-1, 1]$. We take $\delta q_1 = \delta q_2 = \delta p_1 = \delta p_2 = \delta y_1 = \delta y_2 = 1/32$, $N_q = 32$, $N_p = 8$ in FGA, and use $\delta x_1 = \delta x_2 = 1/128$ to reconstruct the solution. Figure 4.5 compares the wave amplitude of the true solution and the one by FGA for $\varepsilon = 1/128$ and $1/256$. The ℓ^∞ and ℓ^2 errors of the wave amplitude are 1.98×10^{-1} and 4.42×10^{-2} for $\varepsilon = 1/128$, and 1.07×10^{-1} and 2.20×10^{-2} for $\varepsilon = 1/256$. This shows a linear convergence in ε of the method.

5. Discussion and conclusion

We first briefly compare the efficiency of frozen Gaussian approximation (FGA) with the Gaussian beam method (GBM). GBM uses only one Gaussian function for each grid point in physical space, while FGA requires more Gaussians per grid point with different initial momentum to capture the behavior of focusing or spreading of the solution. However, the stationary phase approximation suggests that the number of Gaussians is only increased by a small constant multiple of the number of those used in GBM. In addition, in GBM one has to solve the Riccati equation, which is a coupled nonlinear ODE system in high dimension, to get the dynamics of the Hessian matrix for each Gaussian, while in FGA the Hessian matrix is determined initially and has no dynamics. Therefore, the overall efficiency of FGA is comparable to GBM.

Admittedly, higher order GBM gives better asymptotic accuracy, and only requires solving a constant number of additional ODEs as in FGA. The numerical cost of higher order GBM is comparable to FGA. However, higher order GBM has its drawbacks: The imaginary part of higher order (larger than two) tensor function dose not preserve positive definiteness in time evolution, which may destroy the decay property of the ansatz of higher order GBM . This is even more severe when beams spread. Moreover, the ODEs in higher order GBM are in coupled nonlinear systems in high dimension. It raises numerical difficulty caused by stability issues. We also note that the numerical integration of ODEs in FGA can be easily parallelized since the Hamiltonian flow (2.4) is independent for different initial $(\boldsymbol{q}, \boldsymbol{p})$, while it is not so trivial for higher order tensors in GBM.

From the accuracy point of view, our numerical examples show that first order FGA method has asymptotic accuracy $\mathcal{O}(\varepsilon)$. The existing rigorous analysis ([2, 1, 16]) proves that the k-th order GBM has an accuracy of $\mathcal{O}(\varepsilon^{k/2})$. Hence, at the first order, FGA has better asymptotic accuracy than GBM. We note that, however, there has been numerical evidence presenting $\mathcal{O}(\varepsilon)$ asymptotic accuracy order for first order GBM, for example in [9, 11, 18, 16]. This phenomenon is usually attributed to error cancellation between different beams. To the best of our knowledge, the mechanism of error cancellation in GBM has not been systematically understood yet.

With the the gain of half order in asymptotic accuracy due to cancellation, the first order GBM has the same accuracy order as FGA (of course GBM still loses accuracy when beams spread). We should remark that the gain in asymptotic accuracy order

depends on the choice of norm. For example, the first order GBM has a half order convergence in the ℓ^∞ norm, first order convergence in the ℓ^2 norm and 3/2-th order convergence in the ℓ^1 norm in Example 1 of [9]. Moreover, the error cancellation seems not to be easily observed in numerics unless ε is very small. For instance, the convergence order of GBM in Example 4.1 is only a bit better than $1/2$ for ε up to $1/256$. While in FGA, we numerically observe the first order asymptotic accuracy in both ℓ^2 and ℓ^∞ norms.

Actually the accuracy of FGA can also be understood from a viewpoint of error cancellation. Note that the equalities (3.9), (3.10), and (3.11) in Lemma 3.2 play the role of determining the accuracy of FGA. In (3.9), the term $\boldsymbol{x} - \boldsymbol{Q}$ is of order $\mathcal{O}(\sqrt{\varepsilon})$ due to the Gaussian factor, but after integration with respect to $\boldsymbol{q}$ and $\boldsymbol{p}$, which is similar to the beam summation in GBM, it becomes $\mathcal{O}(\varepsilon)$. Similar improvement of order also happens in (3.10) and (3.11). Integration by parts along with (3.14) explains the mechanism of this type of error cancellation.

We conclude the paper as follows. In this work, we propose the frozen Gaussian approximation (FGA) for computation of high frequency wave propagation, motivated by the Herman-Kluk propagator in chemistry literature. This method is based on asymptotic analysis and constructs the solution using Gaussian functions with fixed widths that live on the phase plane. It not only provides an accurate asymptotic solution in the presence of caustics, but also resolves the problem in the Gaussian beam method (GBM) when beams spread. These merits are justified by numerical examples. Additionally, numerical examples also show that FGA exhibits better asymptotic accuracy than GBM. These advantages make FGA quite competitive for computing high frequency wave propagation.

For the purpose of presenting the idea simply and clearly, we only describe the method for the linear scalar wave equation using leading order approximation. The method can be generalized for solving other hyperbolic equations and systems with a character of high frequency. The higher order approximation can also be derived. Since the method is of Lagrangian type, the issue of divergence still remains, which will be resolved in an Eulerian framework. We present these results in the subsequent paper [17].

<div align="center">REFERENCES</div>

[1] S. Bougacha, J.L. Akian, and R. Alexandre, *Gaussian beams summation for the wave equation in a convex domain*, Commun. Math. Sci., 7, 973–1008, 2009.

[2] M. Combescure and D. Robert, *Semiclassical spreading of quantum wave packets and applications near unstable fixed points of the classical flow*, Asympt. Anal., 14, 377–404, 1997.

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

[4] G.B. Folland, *Harmonic Analysis in Phase Space*, Annals of Mathematics Studies, Princeton University Press, Princeton, 122, 1989.

[5] E.J. Heller, *Frozen Gaussians: A very simple semiclassical approximation*, J. Chem. Phys., 75, 2923–2931, 1981.

[6] M.F. Herman and E. Kluk, *A semiclassical justification for the use of non-spreading wavepackets in dynamics calculations*, Chem. Phys., 91, 27–34, 1984.

[7] L. Hörmander, *The Analysis of Linear Partial Differential Operators I, Distribution Theory and Fourier Analysis*, Springer Verlag, New York, 1983.

[8] S. Jin, H. Wu, X. Yang, and Z. Huang, *Bloch decomposition-based Gaussian beam method for the Schrödinger equation with periodic potentials*, J. Comput. Phys., 229, 4869–4883, 2010.

[9] S. Jin, H. Wu, and X. Yang, *Gaussian beam methods for the Schrödinger equation in the semiclassical regime: Lagrangian and Eulerian formulations*, Commun. Math. Sci., 6, 995–1020, 2008.

[10] S. Jin, H. Wu, and X. Yang, *A numerical study of the Gaussian beam methods for one-dimensional Schrödinger-Poisson equations*, J. Comput. Math., 28, 261–272, 2010.

[11] S. Jin, H. Wu, and X. Yang, *Semi-Eulerian and high order Gaussian beam methods for the Schrödinger equation in the semiclassical regime*, Commun. Comput. Phys., 9, 668–687, 2011.

[12] K. Kay, *Integral expressions for the semi-classical time-dependent propagator*, J. Chem. Phys., 100, 4377–4392, 1994.

[13] K. Kay, *The Herman-Kluk approximation: Derivation and semiclassical corrections*, Chem. Phys., 322, 3–12, 2006.

[14] H. Liu and J. Ralston, *Recovery of high frequency wave fields for the acoustic wave equation*, Multiscale Model. Simul., 8, 428–444, 2009.

[15] H. Liu and J. Ralston, *Recovery of high frequency wave fields from phase space based measurements*, Multiscale Model. Simul., 8, 622–644, 2010.

[16] H. Liu, O. Runborg, and N.M. Tanushev, *Error estimates for Gaussian beam superpositions*, arXiv:1008.1320, 2010.

[17] J. Lu and X. Yang, *Frozen Gaussian approximation for general linear strictly hyperbolic system: Formulation and Eulerian methods*, arXiv:1010.1969, 2010.

[18] M. Motamed and O. Runborg, *Taylor expansion and discretization errors in Gaussian beam superposition*, Wave Motion, 47, 421–439, 2010.

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

[20] J. Qian, S. Leung, and R. Burridge, *Eulerian Gaussian beams for high-frequency wave propagation*, Geophysics, 72, 61–76, 2007.

[21] J. Qian and S. Leung, *Eulerian Gaussian beams for Schrödinger equations in the semi-classical regime*, J. Comput. Phys., 228, 2951–2977, 2009.

[22] J. Qian and L.Ying, *Fast Gaussian wavepacket transforms and Gaussian beams for the Schrödinger equation*, J. Comput. Phys., 229, 7848–7873, 2010.

[23] J. Qian and L. Ying, *Fast multiscale Gaussian wavepacket transforms and multiscale Gaussian beams for the wave equation*, Multiscale Model. Simul., to appear.

[24] J. Ralston, *Gaussian beams and the propagation of singularities*, Studies in PDEs, MAA Stud. Math., 23, 206–248, 1982.

[25] D. Robert, *On the Herman-Kluk semiclassical approximation*, Rev. Math. Phys., 22, 1123–1145, 2010.

[26] O. Runborg, *Mathematical models and numerical methods for high frequency waves*, Commun. Comput. Phys., 2, 827–880, 2007.

[27] T. Swart and V. Rousse, *A mathematical justification of the Herman-Kluk propagator*, Commun. Math. Phys., 286, 725–750, 2009.

[28] N.M. Tanushev, J. Qian, and J. Ralston, *Mountain waves and Gaussian beams*, Multiscale Model. Simul., 6, 688–709, 2007.

[29] N.M. Tanushev, *Superpositions and higher order Gaussian beams*, Commun. Math. Sci., 6, 449–475, 2008.