The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific
and Cultural Organization

International Atomic
Energy Agency

SMR.1803 - 1

SCHOOL ON MATHEMATICAL METHODS FOR OPTICS

23 - 27 January 2006

Asyptotic Methods in Optics:
Using rays to describe waves

Miguel A. ALONSO
The Institute of Optics
University of Rochester
Rochester NY 14607
U.S.A.

Asymptotic methods in optics: Using rays to describe waves

Miguel A. Alonso

The Institute of Optics,

University of Rochester,

Rochester NY 14607, U.S.A.

School on Mathematical Methods in Optics

Abdus Salam International Centre for Theoretical Physics

January 23-27 2006

I. INTRODUCTION

In optics, we often need to model the propagation of a wave field through a complex medium. However, except in trivial cases, the propagation of the field cannot be described analytically, so it is necessary to use approximations and/or numerical schemes. Some of these schemes are based on the ray model (geometrical optics). It could be argued, however, that the ray model is a primitive and "outdated" description of light, which does not account for wave effects like interference and diffraction, and therefore has limited applicability. Why do we keep using it so much then? Because it is very simple and gives a sufficiently good description of the performance of optical systems for many practical purposes. For example, the design of most large optical instruments is essentially based on geometrical optics. This situation is analogous to that of mechanical design: bridges, tools, cars, etc. are always designed and described using classical mechanics, which is also an "outdated" theory. A mechanical engineer would never contemplate using rigorous quantum mechanics in the design of a macroscopic instrument! Quantum effects are important only for very small (or very special) systems. It turns out that, even in these situations, an accurate description can be achieved using classical mechanics. As we will see later, this analogy between the classical/quantum models in mechanics and the ray/wave models in optics is of a deep mathematical nature.

The objective of this course is to learn how the ray model can actually be used in wave calculations. This will also give us some insight into why geometrical optics works so well, and under what conditions.

In what follows, we will consider for simplicity the scalar model. We will also limit ourselves to the case of monochromatic light, where the field's time dependence can be factored as $E(\vec{r}, t) = U(\vec{r}) \exp(-i\omega t)$. Of course, polychromatic light can be made up of monochromatic components. The basic equation that describes the propagation of a monochromatic scalar field $U(\vec{r})$, where $\vec{r} = (x, y, z)$ is the position vector, is the Helmholtz equation

$$[\nabla^2 + k^2 n^2(\vec{r})]U(\vec{r}) = 0, \tag{1}$$

where $k = \omega/c$ is the wavenumber (with ω being the frequency, c the speed of light in vacuum), and $n(\vec{r})$ is the position-dependent refractive index. It is assumed here that this refractive index is real (that is, that the medium presents no absorption or gain) and varies smoothly. The case of abrupt interfaces will be discussed later.

II. ASYMPTOTICS IN THE POSITION REPRESENTATION

In order to find the connection between wave optics and geometrical optics, let us assume that the field U consists of a slowly-varying amplitude and a rapidly-oscillating phase proportional to the wavenumber[1, 2], i.e.

$$U(\vec{r}) = A(\vec{r}) \exp[ik\phi(\vec{r})], \tag{2}$$

where it is assumed that at least ϕ is real. The substitution of Eq. (2) into Eq.(1) gives, after some reordering and multiplication by $\exp(-ik\phi)$,

$$k^2 A(n^2 - \nabla\phi \cdot \nabla\phi) + ik(2\nabla A \cdot \nabla\phi + A\nabla^2\phi) + \nabla^2 A = 0. \tag{3}$$

The goal is now to separate this equation into two (or more) equations that are amenable to simpler solution or that at least have an interesting interpretation. There are (at least) two ways to proceed:

A) Assume that the wavenumber k is very large, and use an asymptotic treatment.

B) Assume that both A and ϕ are real, and separate Eq. (3) into real and imaginary parts.

We will explore both these possibilities. Let us start with A), since it is the approach we are more interested in, and because knowing A) will help us interpret B).

A. Asymptotic treatment

1. *The top order: Geometrical optics*

Since k is very large, the leading part of Eq. (3) is the one proportional to k^2. If we want to make the whole equation equal to zero, it is convenient to make that part zero independently first. This means that

$$\nabla\phi(\vec{r}) \cdot \nabla\phi(\vec{r}) = n^2(\vec{r}). \tag{4}$$

This equation is the well-known eikonal equation, which as we will show gives rise to geometrical optics. The function ϕ is called the eikonal function or simply the eikonal.

The eikonal equation can be solved (or at least written in a form better suited for numerical computation) by parametrizing the position vector in terms of three independent parameters τ, ξ_1, ξ_2 as $\vec{r} = \vec{R}(\tau, \xi_1, \xi_2)$. These parameter must be chosen such that $\vec{R}$ moves in three different directions as we vary each of them, i.e., the three vectors corresponding to the partial derivatives of $\vec{R}$ with respect to each of the parameters are linearly independent. In order to solve the eikonal equation, we choose the partial derivative of $\vec{R}$ with respect to τ to be parallel to the gradient of the eikonal, i.e.

$$\frac{\partial \vec{R}}{\partial \tau} = \dot{\vec{R}} = \alpha \nabla \phi(\vec{R}), \tag{5}$$

where the proportionality function $\alpha(\tau, \xi_1, \xi_2)$ is assumed to be positive. From substituting the gradient of the eikonal as given in Eq. (5) into Eq. (4) evaluated at $\vec{R}$, we find

$$\alpha(\tau, \xi_1, \xi_2) = \frac{h(\tau, \xi_1, \xi_2)}{n[\vec{R}(\tau, \xi_1, \xi_2)]}, \tag{6}$$

where

$$h(\tau, \xi_1, \xi_2) := |\dot{\vec{R}}(\tau, \xi_1, \xi_2)|, \tag{7}$$

with ":=" denoting a definition.

To find the equations that determine the evolution of the rays, it is convenient to define the "momentum" vector $\vec{P}$ as

$$\vec{P}(\tau, \xi_1, \xi_2) := \nabla \phi[\vec{R}(\tau, \xi_1, \xi_2)] = \frac{n(\vec{R})}{h} \dot{\vec{R}}, \tag{8}$$

where Eqs. (5) and (6) were used in the last step. The equation for the evolution of the rays is found by considering the derivative with respect to τ of the first two parts of Eq. (8):

$$\dot{\vec{P}} = (\dot{\vec{R}} \cdot \nabla)\nabla\phi = \frac{h}{n}(\nabla\phi \cdot \nabla)\nabla\phi = \frac{h}{n}\frac{1}{2}\nabla(\nabla\phi \cdot \nabla\phi) = \frac{h}{n}\frac{1}{2}\nabla(n^2) = h\nabla n(\vec{R}), \tag{9}$$

where the chain rule was used in the first step, the second part of Eq. (8) was used in the second step, and the eikonal equation was used in the third step. Finally, to obtain the phase, let us define

$$L(\tau, \xi_1, \xi_2) := \phi[\vec{R}(\tau, \xi_1, \xi_2)]. \tag{10}$$

This function increases with τ according to the expression

$$\dot{L} = \nabla\phi \cdot \dot{\vec{R}} = n(\vec{R})h, \tag{11}$$

where the chain rule, as well as Eqs. (8) and (7), were used. The value of the phase is then found by integrating this expression in τ.

The trajectories traced by the vector $\vec{R}$ for increasing τ are precisely the rays of geometrical optics. The expressions in Eqs. (8), (9), and (11) are the basic equations that rule the propagation of the rays. Let's summarize these equations:

$$\dot{\vec{R}} = \frac{h}{n(\vec{R})}\vec{P}, \tag{12}$$

$$\dot{\vec{P}} = h\,\nabla n(\vec{R}), \tag{13}$$

$$\dot{L} = h\,n(\vec{R}), \tag{14}$$

where, recall, $h = |\dot{\vec{R}}|$ is the "speed" of the parametrization. The first two of these equations (which are coupled) rule the propagation of the ray, while the third one describes how the eikonal or "optical path length" increases under propagation. Strictly speaking, Eq. (12) is just the geometrical definition of the optical momentum as a vector which is locally tangent to the ray, and whose magnitude is given by the local refractive index. Equation (13), on the other hand, tells us that local changes in the refractive index change the direction of propagation of the ray. When one considers the interface between two homogeneous media with different refractive indices, this equation gives Snell's law. When the refractive index changes continuously and smoothly, the rays change direction gradually and become curved.

While varying the parameter τ causes $\vec{R}$ to move along a ray, variations of ξ_1 or ξ_2 make $\vec{R}$ jump to another ray. That is, the above set of equations describes the evolution of a two-parameter family of rays, where each ray corresponds to a set of values of the parameters $\boldsymbol{\xi} = (\xi_1, \xi_2)$. The evolution of each ray is completely autonomous. This can be appreciated from Eqs. (12) and (13), which involve no operation on the parameters ξ_1, ξ_2. The optical path length of the different rays in the family are interconnected, nevertheless, as one can see from considering the derivative of Eq. (10) with respect to one of these parameters:

$$\frac{\partial L}{\partial \xi_j} = \nabla\phi \cdot \frac{\partial \vec{R}}{\partial \xi_j} = \vec{P} \cdot \frac{\partial \vec{R}}{\partial \xi_j}. \tag{15}$$

This equation tells us that the surfaces of constant L are perpendicular to the rays.

2. Choosing z as the parameter

The equations for the rays found above are general in the sense that the parametrization along the rays is arbitrary. There are, however, particular parametrizations that are convenient and that have been studied in the past. These parametrizations make the ray equations take several different forms. Some common choices are the ones that make τ be equal to: the arc-length of the ray (so $h = 1$), the optical path length (so $h = 1/n$), or the length divided by the local refractive index (so $h = n$). In what follows, we will concentrate on a fourth particular parametrization which, while being more limited in application, is convenient for the type of ideas we will study. This parametrization is only valid when one can choose a "main direction of propagation" such that the component of the momentum in this direction for all rays in the family is always positive. Let us align the z axis with this direction of propagation. The condition for the application of this parametrization is then that the rays do not "turn around" in z, so that their positions are single valued functions of z. Under these circumstances, z itself can be used as the parameter of propagation.

It is now convenient to separate the z components from the transverse components of the position and momentum vectors as

$$\vec{R}(z,\xi) = [X(z,\xi), Y(z,\xi), z] = [\mathbf{X}(z,\xi), z], \tag{16}$$

$$\vec{P}(z,\xi) = [P_x(z,\xi), P_y(z,\xi), H(z,\xi)] = [\mathbf{P}(z,\xi), H(z,\xi)], \tag{17}$$

where $\mathbf{X} = (X, Y)$ and $\mathbf{P} = (P_x, P_y)$. As stated earlier, it is assumed that the longitudinal component of the momentum, namely H, is always positive. As we know from the eikonal equation (4), this function is related to $\mathbf{P}$ and the refractive index by

$$H(z,\xi) = \sqrt{n^2(\mathbf{X}, z) - \mathbf{P}\cdot\mathbf{P}}. \tag{18}$$

Let us now find the form that the ray equations take for this parametrization. We start by considering the longitudinal (i.e. z) part of Eq. (12):

$$1 = \frac{hH}{n}. \tag{19}$$

From this expression we find that $h = n/H$. The transverse part of Eq. (12), on the other hand, gives

$$\dot{\mathbf{X}} = \frac{h}{n}\mathbf{P} = \frac{\mathbf{P}}{H}. \tag{20}$$

Similarly, the longitudinal and transverse parts of Eq. (13) become, respectively

$$\dot{H} = \frac{n}{H}\frac{\partial n}{\partial z}(\mathbf{X}, z) = \frac{1}{2H}\frac{\partial n^2}{\partial z}(\mathbf{X}, z), \tag{21}$$

$$\dot{\mathbf{P}} = \frac{n}{H}\frac{\partial n}{\partial \mathbf{x}}(\mathbf{X}, z) = \frac{1}{2H}\frac{\partial n^2}{\partial \mathbf{x}}(\mathbf{X}, z), \tag{22}$$

where $\partial/\partial\mathbf{x}$ is the transverse (i.e. x, y) part of the gradient. Finally, Eq. (14) becomes

$$\dot{L} = \frac{n^2(\mathbf{X}, z)}{H}. \tag{23}$$

3. The rest of the orders: the transport equation

After we chose ϕ to satisfy the eikonal equation, Eq. (3) can be written as

$$2\nabla A \cdot \nabla\phi + A\nabla^2\phi + \frac{1}{ik}\nabla^2 A = 0. \tag{24}$$

The key of asymptotic treatments like this is to regard k as a variable (even though it is a constant!) which takes very large values. We then write A as a so-called Debye series of the form

$$A(\vec{r}) = \sum_{j=0}^{\infty}\frac{A_j(\vec{r})}{(ik)^j}. \tag{25}$$

This series is substituted into Eq. (24) and rearranged to get

$$2\nabla A_0 \cdot \nabla\phi + A_0\nabla^2\phi + \sum_{j=1}^{\infty}\frac{1}{(ik)^j}(2\nabla A_j \cdot \nabla\phi + A_j\nabla^2\phi + \nabla^2 A_{j-1}) = 0. \tag{26}$$

The coefficient of each power of k is now forced to be made zero independently. This gives a hierarchy of linked equations for each of the A_j's in the form:

$$2\nabla A_0 \cdot \nabla\phi + A_0\nabla^2\phi = 0, \tag{27}$$

$$2\nabla A_j \cdot \nabla\phi + A_j\nabla^2\phi = -\nabla^2 A_{j-1}, \quad j = 1, 2, ... \tag{28}$$

Equation (27) can be solved to find A_0, and the rest of the A_j's can be found successively in terms of the previous one by solving Eq. (28). For sufficiently large k, however, $A \sim A_0$, so we will only care about Eq. (27). Notice that, by multiplying both sides of this equation by A_0, it can be rewritten as

$$\nabla \cdot (A_0^2\nabla\phi) = 0. \tag{29}$$

This expression is known as the transport equation. In order to solve it, consider integrating both sides over the volume occupied by an infinitesimally thin bundle of rays $\mathcal{B}$ corresponding to small intervals $\Delta\xi_1, \Delta\xi_2$ around a central ray ξ_1, ξ_2, and to z between z_0 and z_1. By using Gauss's theorem, we obtain

$$\int_{\mathcal{B}} \nabla \cdot (A_0^2 \nabla\phi)\, d^3r = \int_{\partial\mathcal{B}} A_0^2 \nabla\phi \cdot d\vec{a} = 0, \tag{30}$$

where $\partial\mathcal{B}$ refers to the outer surface of the bundle $\mathcal{B}$, and $d\vec{a}$ is an outward-pointing area element. It is easy to see that the only contributions to the surface integral come from the end faces of the bundle, as the area element normal is perpendicular to the ray momentum $\nabla\phi$ at the sides of the bundle. Let the infinitesimally small area elements at both ends of the bundle be called Δa_0 and Δa_1, respectively, so Eq. (30) can be written as

$$A_0^2[\mathbf{X}(\xi, z_0), z_0] H(\xi, z_0)\, (-\Delta a_0) + A_0^2[\mathbf{X}(\xi, z_1), z_1] H(\xi, z_1)\, \Delta a_1 = 0, \tag{31}$$

where the minus sign in the area element for the first term comes from the fact that $\nabla\phi$ points into $\mathcal{B}$ at the beginning of the bundle, while it points out of $\mathcal{B}$ at the end of the bundle. In getting this expression, we also used the fact that the z component of $\nabla\phi[\mathbf{X}(\xi, z), z]$ is simply H. The intensity of the field is given by $|A|^2 \sim A_0^2$. The product of this intensity and H (which is the refractive index times an obliquity factor) is proportional to the flux density traversing the area element. Therefore Eq. (31) has an intuitive interpretation: the total flux entering the bundle at one end equals the flux exiting at the other end. This means that the rays behave as infinitesimal conduits of power. When the bundle expands, causing the area at the exit of the bundle to be bigger than the area at the entrance, the flux density (and therefore the intensity) of the field becomes smaller in the same proportion, since the power is spreading over a larger area.

Now, notice that the area elements are given by

$$\Delta a_j = \left.\frac{\delta(\mathbf{X})}{\delta(\xi)}\right|_{z_j} \Delta\xi_1 \Delta\xi_2 = \left.\left(\frac{\partial X}{\partial \xi_1}\frac{\partial Y}{\partial \xi_2} - \frac{\partial X}{\partial \xi_2}\frac{\partial Y}{\partial \xi_1}\right)\right|_{z_j} \Delta\xi_1 \Delta\xi_2. \tag{32}$$

By replacing z_1 with z, we can solve Eq. (31) for $A_0(\mathbf{X}, z)$ which, after using Eq. (32), gives

$$\mathcal{A}_0(z, \xi) := A_0[\mathbf{X}(z, \xi), z] = \sqrt{\frac{H(\xi, z_0)}{H(\xi, z)} \left.\frac{\delta(\mathbf{X})}{\delta(\xi)}\right|_{z_0} \Big/ \left.\frac{\delta(\mathbf{X})}{\delta(\xi)}\right|_{z}}\; A_0[\mathbf{X}(z_0, \xi), z_0]. \tag{33}$$

This expression is the final piece that we needed to estimate the field.

4. The field estimate and its problems

We now will put all the pieces together for the construction of the field estimate. The estimate of the solution of the Helmholtz equation is found in terms of the parametric equation

$$U[\mathbf{X}(z,\boldsymbol{\xi}),z] \sim \mathcal{A}_0(z,\boldsymbol{\xi})\,\exp[ikL(z,\boldsymbol{\xi})]. \tag{34}$$

The approximation results from neglecting all terms but the first in the Debye series for A, which is justified by the fact that k is large.

To use the formula in Eq. (34), one must determine somehow from the initial conditions of the field, initial conditions for the rays and the amplitude. Once this is done, the rays can be traced by using the ray equations, and from there the amplitude and phase of the field can be computed. A remarkable aspect of this formula is that the field information seems to be transmitted through the rays: the value of the field at any point is determined only by the infinitesimal bundle of rays surrounding the ray that goes through this point. That is, if we only integrate the ray equations over a thin bundle of rays, we can estimate the field along this bundle over a long distance, regardless of what the rest of the rays do.

Notice, however, that the expression for the amplitude of this estimate given in Eq. (33) diverges when either of the two following conditions happens:

$$H(z,\boldsymbol{\xi}) = 0, \tag{35}$$

$$\left.\frac{\delta(\mathbf{X})}{\delta(\boldsymbol{\xi})}\right|_z = 0. \tag{36}$$

The first of these conditions happens when rays turn around in z. As we mentioned at the outset, we are assuming that such a situation will not happen. The second condition, on the other hand, happens when the bundle collapses, i.e. when the rays that make up the bundle cross. This crossing of contiguous rays is what is known as a caustic. While the field at a caustic is indeed large, it is not infinite as Eq. (33) predicts. This signals a problem in this formalism. Near the caustic, the field estimates become unrealistically large. The problem comes from the fact that, in the vicinity of the caustic, the left-hand side of Eq. (28) becomes large due to the fast variation of A_0. This causes the A_js to be large too, up to a point where the Debye series cannot be approximated by its leading term. The field estimate in Eq. (34) therefore breaks down in the vicinity of caustics.

Remarkably, though, after the caustic, if the rays become sufficiently separated and uniform, this field estimate becomes valid again. It must be noted, however, that the passage through the caustic gives rise to changes in sign for the Jacobian inside the square root in Eq. (33). Due to the square root, these sign changes cause phase shifts which are integer multiples of $\pi/2$, where the integer is called the Maslov or Morse index. The determination of these indices can be quite complicated, as the appropriate sign of the square root must be chosen. (Note: one particular case of this phase shift is the Gouy phase shift undergone by a focused beam.) In many situations, after a caustic or caustics, there are several bundles of rays occupying the same volume. The field estimate then results from summing their contributions, each with a suitable Maslov phase. (This is one example of how geometrical optics can account for interference effects.) Also, there might be regions to one side of a caustic where no rays arrive. The simple estimate presented above then offers no access to the field in this region, and one assumes that the field is simply zero (although this is not strictly the case). There have been variations of the scheme presented here which give estimates in these dark regions by using the so-called complex rays, i.e. rays that go through complex positions. This more complicated approach is beyond the scope of this course. Some of the methods that will be described in what follows, however, do not present this problem.

B. The Bohmian approach

Let us go back to the alternative B) mentioned towards the beginning of this section, where it was proposed to assume that both A and ϕ are real, so Eq. (3) can be separated into two equations, corresponding to the real and the imaginary parts. These two equations can be written as

$$\nabla\phi\cdot\nabla\phi = n^2 - \frac{\nabla^2 A}{k^2 A}, \tag{37}$$

$$\nabla\cdot(A^2\nabla\phi) = 0. \tag{38}$$

Notice that these equations look very similar to those obtained earlier. In particular, Eq. (38) is identical to Eq. (29), except for the fact that we have the full amplitude A instead of A_0, which was only the leading term in a series. That is, here it is not necessary to express A as a Debye series, and the full A can in principle be found by integrating Eq. (38) over an infinitesimal bundle of special trajectories. These trajectories must be found by solving Eq. (37) parametrically. Notice that this expression looks a lot like the eikonal equation in Eq. (4), but where a correction term has been added. We will call this term the Bohmian correction for reasons that will be explained later. This correction includes A, so the two equations above are coupled: one can no longer first solve the eikonal equation to find the rays and then use these rays to solve the transport equation. The Bohmian correction is proportional to k^{-2}, so for large k this correction is usually very small, meaning that the corresponding trajectories or " Bohmian rays" are very similar to the standard rays almost everywhere. The Bohmian correction has a similar effect to that of the refractive index: its variation causes the rays to bend. This extra bending is negligible except in regions where A varies quickly, e.g. near a focus. The effect of the Bohmian correction is quite remarcable: it makes the rays act as if they repelled each other, turning away from each other instead of crossing. Since Bohmian rays never cross, there is always only one Bohmian ray at any point in space. This seems like a good thing, because the problems with caustics disappear. However, this approach has two serious disadvantages. First, the fact that the eikonal and the transport equation are coupled greatly complicates the determination of the rays, which must be found numerically even in very simple cases like propagation in free space. Second, the method has problems at zeros of the field, since the Bohmian potential can blow up when $A = 0$.

C. Quantum mechanics

All of the ideas presented so far can also be applied to the study of quantum-mechanical problems[3]. Let us concentrate in the case of non-relativistic quantum mechanics for a single particle of mass m. This problem is ruled by the Schrödinger equation.

$$i\hbar\frac{\partial\Psi}{\partial t}(\vec{r},t) = -\frac{\hbar^2}{2m}\nabla^2\Psi(\vec{r},t) + V(\vec{r},t)\Psi(\vec{r},t), \tag{39}$$

where $\hbar$ is the reduced Planck constant and V is the potential. The wavefunction Ψ plays the role of the wavefield. Just as with the Helmholtz equation, let us write the wave function in terms of a slowly-varying amplitude and a rapidly oscillating phase, i.e.

$$\Psi(\vec{r},t) = A(\vec{r},t)\,\exp\left[\frac{i}{\hbar}\phi(\vec{r},t)\right]. \tag{40}$$

Notice that we made the phase proportional to the inverse of $\hbar$. Therefore, like k for the Helmholtz case, $\hbar^{-1}$ will play the role of the large asymptotic parameter. The approximate results we will find will then be valid in the so-called semiclassical regime, i.e. when $\hbar$ is small compared to all the quantities (or variations of quantities) that present the same units (action=position$\times$momentum).

After substituting of Eq. (40) and reordering, Eq. (39) can be written as

$$\left(\frac{\partial\phi}{\partial t}+\frac{\nabla\phi\cdot\nabla\phi}{2m}+V\right)-i\hbar\left(\frac{\partial A}{\partial t}+\frac{2\nabla A\cdot\nabla\phi+A\nabla^2\phi}{2m}\right)-\hbar^2\,\nabla^2 A=0. \tag{41}$$

Again, we face a decision between two approaches:

A) Assume that $\hbar$ is very small, and expand A in a Debye series.

B) Assume that both A and ϕ are real, and separate the real and imaginary parts of Eq. (41).

As before, let us consider the approach A) first.

1. *Semiclassical mechanics*

We start by making the dominant term in Eq. (41) equal zero, that is:

$$\frac{\partial\phi}{\partial t}+\frac{\nabla\phi\cdot\nabla\phi}{2m}+V=0. \tag{42}$$

This is the equation for the action in classical mechanics. It can be solved parametrically in a fashion similar to the eikonal equation. We start by parametrizing the position as $\vec{r}=\vec{R}(t,\xi_1,\xi_2,\xi_3)$. We then define the momentum $\vec{\mathcal{P}}$ and the Hamiltonian as $\mathcal{H}$, i.e.

$$\vec{\mathcal{P}}:=\nabla\phi(\vec{R},t),\;\;\mathcal{H}:=-\frac{\partial\phi}{\partial t}(\vec{R},t), \tag{43}$$

and we choose the momentum to be proportional to the time derivative of the position, i.e

$$\vec{\mathcal{P}}=m\dot{\vec{R}}. \tag{44}$$

The equation for the evolution of the "rays" for this problem results from considering the time derivative of the first equation in Eq. (43):

$$\dot{\vec{\mathcal{P}}}=(\dot{\vec{R}}\cdot\nabla)\nabla\phi+\nabla\frac{\partial\phi}{\partial t}=\left(\frac{\nabla\phi}{m}\cdot\nabla\right)\nabla\phi+\nabla\frac{\partial\phi}{\partial t}=\nabla\left(\frac{\nabla\phi\cdot\nabla\phi}{2m}+\frac{\partial\phi}{\partial t}\right)=-\nabla V(\vec{R},t). \tag{45}$$

This equation is clearly Newton's second law of motion, so the "rays" of the Schrödinger equation are classical trajectories. In other words, the mathematical relation between wave and ray optics is the same as that between quantum and classical mechanics. The phase function, i.e. the action, can also be parametrized as $\mathcal{S}:=\phi(\vec{R},t)$, and its equation of evolution results from considering

$$\dot{\mathcal{S}}=\dot{\vec{R}}\cdot\nabla\phi+\frac{\partial\phi}{\partial t}=\frac{\vec{\mathcal{P}}\cdot\vec{\mathcal{P}}}{2m}-V(\vec{R},t). \tag{46}$$

Notice that the right-hand side of this expression is the Lagrangian.

Let us now consider the rest of Eq. (41). As mentioned earlier, we propose a Debye expansion for the amplitude in the form

$$A(\vec{r},t) = \sum_{j=0}^{\infty} (i\hbar)^j A_j(\vec{r},t). \tag{47}$$

After substituting this into Eq. (41) (recall that the first term has been made to vanish), separating the different powers of $\hbar$ and rearranging, we get the equations

$$\frac{\partial A_0^2}{\partial t} + \nabla \cdot \left(A_0^2 \frac{\nabla \phi}{m} \right) = 0, \tag{48}$$

$$\frac{\partial A_j}{\partial t} + \frac{2\nabla A_j \cdot \nabla \phi + A_j \nabla^2 \phi}{2m} = -\nabla^2 A_{j-1}, \tag{49}$$

fore $j \geq 1$. Notice that Eq. (48) is a continuity equation. Just like in the optical case, it can be solved to give

$$\mathcal{A}_0 = A_0[\vec{R}(t,\xi_1,\xi_2,\xi_3),t] = \sqrt{\left.\frac{\delta(\vec{R})}{\delta(\xi_1,\xi_2,\xi_3)}\right|_{t_0} \Big/ \left.\frac{\delta(\vec{R})}{\delta(\xi_1,\xi_2,\xi_3)}\right|_{t}}\, A_0[\vec{R}(t_0,\xi_1,\xi_2,\xi_3),t_0]. \tag{50}$$

These results are the basis of many semiclassical techniques used to understand and/or calculate quantum-mechanical results based on classical mechanics. One such technique is the well-known WKB (or JWKB) method[4] and the Van Vleck-Gutzwiller propagator[5, 6]. However, the amplitude estimate in Eq. (50) diverges when classical trajectories cross. This happens, for example, at the classical turning points when solving a time-independent problem. This problem is completely analogous to the caustic problem in optics.

2. *Bohmian mechanics and the hydrodynamic model*

Now let us consider the other approach, where A is assumed to be real. After a bit of simple manipulation, the real and imaginary parts of Eq. (41) can be written as

$$\frac{\partial \phi}{\partial t} + \frac{\nabla \phi \cdot \nabla \phi}{2m} + V - \hbar^2 \frac{\nabla^2 A}{A} = 0, \tag{51}$$

$$\frac{\partial A^2}{\partial t} + \nabla \cdot \left(A^2 \frac{\nabla \phi}{m} \right) = 0, \tag{52}$$

This form of separating Schrödinger's equation is the basis of David Bohm's (and Louis deBroglie's) pilot wave interpretation for quantum mechanics[7]. Notice that Eq. (51) is almost identical to the classical equation for the action, except for the extra term $-\hbar^2\nabla^2 A/A$. This term is referred to as the quantum potential, and as what we called the Bohmian term in the optical case, it has the effect of modifying the classical motion of the particles and keeping them from crossing. The interpretation of deBroglie and Bohm is that there is a directly undetectable "guiding wave" whose behavior is ruled by Schrödinger's equation. This wave modifies the classical motion of the particles, which behave then as "visible" surfers riding an "invisible" wave. This interpretation differs from the so-called Copenhagen interpretation of quantum mechanics, where the wavefunction is a complex amplitude whose squared modulus is a probability density.

Besides the philosophical interpretation of these results, the expressions above serve as the basis for computational methods. This formalism is referred to as the hydrodynamic model since, as seen from Eq. (52), the square modulus of the wavefunction satisfies a continuity equation akin to that of an incompressible fluid. However, as in the optical case, the fact that Eqs. (51) and (52) are coupled makes their solution difficult, either algebraically or computationally, especially when the wavefunction presents zeros.

III. ASYMPTOTICS IN THE MOMENTUM REPRESENTATION

The asymptotic approach presented in Section II yielded field estimates where the field at a point is only due to the rays that go through that point. That is, rays are completely local entities. This approach lead to problems at caustics, where there is an infinite density of rays locally going through the point. In this section we will explore an alternative asymptotic approach, where instead of working with the field, we will use its Fourier transform over the transverse coordinates, defined as

$$\tilde{U}(p_x, p_y, z) = \frac{k}{2\pi} \iint_{-\infty}^{\infty} U(x, y, z)\, \exp(-ik\mathbf{x} \cdot \mathbf{p})\, dxdy, \tag{53}$$

where $\mathbf{x} = (x, y)$ and $\mathbf{p} = (p_x, p_y)$ (Note: here we are choosing to transform over all transverse coordinates. It is also possible to only transform over one of them.) In free space, this Fourier transform is what is known as the angular spectrum. In a smoothly inhomogeneous refractive medium, $\tilde{U}$ satisfies an equation corresponding to the Fourier transformation of both sides of Eq. (1):

$$\left[-k^2\mathbf{p} \cdot \mathbf{p} + \frac{\partial^2}{\partial z^2} + k^2 n^2 \left(\frac{i}{k} \frac{\partial}{\partial \mathbf{p}}, z \right) \right] \tilde{U} = 0, \tag{54}$$

where

$$n^2\left(\frac{i}{k}\frac{\partial}{\partial \mathbf{p}}, z\right)\tilde{U} = \sum_{j=0}^{\infty}\frac{1}{j!}\left(\frac{i}{k}\right)^j \left[n^2(\mathbf{x}, z)\left(\frac{\overleftarrow{\partial}}{\partial \mathbf{x}} \cdot \frac{\overrightarrow{\partial}}{\partial \mathbf{p}}\right)^j \tilde{U}(\mathbf{p}, z)\right]\Bigg|_{\mathbf{x}=(0,0)}, \tag{55}$$

where the arrows indicate the direction in which the derivatives act.

In order to perform the asymptotic treatment, we now write $\tilde{U}$ as

$$\tilde{U} = B(\mathbf{p}, z)\, \exp[ik\Gamma(\mathbf{p}, z)]. \tag{56}$$

Notice that the substitution of this form in Eq. (55) gives

$$\begin{aligned}
n^2\left(\frac{i}{k}\frac{\partial}{\partial \mathbf{p}}, z\right)[B\, \exp(ik\Gamma)] = \sum_{j=0}^{\infty}\left(\frac{i}{k}\right)^j n^2(\mathbf{x}, z)\Bigg[&\frac{1}{j!}\left(ik\frac{\overleftarrow{\partial}}{\partial \mathbf{x}} \cdot \frac{\partial \Gamma}{\partial \mathbf{p}}\right)^j B \\
&+ \frac{1}{(j-1)!}\left(ik\frac{\overleftarrow{\partial}}{\partial \mathbf{x}} \cdot \frac{\partial \Gamma}{\partial \mathbf{p}}\right)^{j-1}\left(ik\frac{\overleftarrow{\partial}}{\partial \mathbf{x}} \cdot \frac{\partial B}{\partial \mathbf{p}}\right) \\
&+ \frac{1}{2(j-2)!}\left(ik\frac{\overleftarrow{\partial}}{\partial \mathbf{x}} \cdot \frac{\partial \Gamma}{\partial \mathbf{p}}\right)^{j-2}\left(ik\frac{\overleftarrow{\partial}}{\partial \mathbf{x}} \cdot \frac{\partial^2 B}{\partial \mathbf{p}\partial \mathbf{p}} \cdot \frac{\overleftarrow{\partial}}{\partial \mathbf{x}}\right) \\
&+ \mathcal{O}(k^{-2})\Bigg]\Bigg|_{\mathbf{x}=(0,0)} \exp(ik\Gamma) \\
= \Bigg[&Bn^2\left(-\frac{\partial \Gamma}{\partial \mathbf{p}}, z\right) + \frac{i}{k}\frac{\partial B}{\partial \mathbf{p}} \cdot \frac{\partial n^2}{\partial \mathbf{x}}\left(-\frac{\partial \Gamma}{\partial \mathbf{p}}, z\right) \\
&- \frac{iB}{2k}\mathrm{Tr}\left\{\frac{\partial^2 n^2}{\partial \mathbf{x}\partial \mathbf{x}}\left(-\frac{\partial \Gamma}{\partial \mathbf{p}}, z\right) \cdot \frac{\partial^2 \Gamma}{\partial \mathbf{p}\partial \mathbf{p}}\right\} + \mathcal{O}(k^{-2})\Bigg] \exp(ik\Gamma),
\end{aligned} \tag{57}$$

where we only wrote explicitly the two leading orders in powers of k. With this, Eq. (54) can be written, after dividing by $-k^2 \exp(ik\Gamma)$, as

$$\begin{aligned}
&B\left[\mathbf{p} \cdot \mathbf{p} + \left(\frac{\partial \Gamma}{\partial z}\right)^2 - n^2\left(-\frac{\partial \Gamma}{\partial \mathbf{p}}, z\right)\right] \\
&+ \frac{1}{ik}\left[2\frac{\partial B}{\partial z}\frac{\partial \Gamma}{\partial z} + B\frac{\partial^2 \Gamma}{\partial z^2} + \frac{\partial B}{\partial \mathbf{p}} \cdot \frac{\partial n^2}{\partial \mathbf{x}}\left(-\frac{\partial \Gamma}{\partial \mathbf{p}}, z\right) - \frac{1}{2}\mathrm{Tr}\left\{\frac{\partial^2 n^2}{\partial \mathbf{x}\partial \mathbf{x}}\left(-\frac{\partial \Gamma}{\partial \mathbf{p}}, z\right) \cdot \frac{\partial^2 \Gamma}{\partial \mathbf{p}\partial \mathbf{p}}\right\}\right] \\
&+ \mathcal{O}(k^{-2}) = 0.
\end{aligned} \tag{58}$$

We can again use the same two approaches mentioned earlier to separate Eq. (58) into different parts that are easier to solve. For example, in principle we could assume that B is real and separate this equation into real and imaginary parts. This would lead to a "momentum Bohmian" formalism where no two rays ever have the same momentum at a given z although they can cross freely in position. However, notice that the two resulting equations would be very complicated (except for very simple refractive index distributions), involving terms of many orders in k. The asymptotic treatment, on the other hand, is tractable. We start by expanding B in a Debye series, i.e.

$$B(\mathbf{p}, z) = \sum_{j=0}^{\infty} \frac{B_j}{(ik)^j}. \tag{59}$$

The substitution of this series into Eq. (58) gives analogs of the eikonal and transport equations

$$\mathbf{p} \cdot \mathbf{p} + \left(\frac{\partial \Gamma}{\partial z}\right)^2 = n^2\left(-\frac{\partial \Gamma}{\partial \mathbf{p}}, z\right), \tag{60}$$

$$\frac{\partial}{\partial z}\left(B_0^2 \frac{\partial \Gamma}{\partial z}\right) + \frac{1}{2}\frac{\partial}{\partial \mathbf{p}} \cdot \left(B_0^2 \frac{\partial \Gamma}{\partial \mathbf{p}}\right) = 0. \tag{61}$$

Equation (60) can be solved by parametrizing $\mathbf{p} = \mathbf{P}(z, \xi)$ and defining

$$\mathbf{X}(z, \xi) := -\frac{\partial \Gamma}{\partial \mathbf{p}}(\mathbf{P}, z), \tag{62}$$

$$H(z, \xi) := \frac{\partial \Gamma}{\partial z}(\mathbf{P}, z). \tag{63}$$

Notice, that the derivative with respect to z of both sides of Eq. (62) gives

$$\dot{\mathbf{X}} = -\left(\dot{\mathbf{P}} \cdot \frac{\partial}{\partial \mathbf{p}}\right)\frac{\partial \Gamma}{\partial \mathbf{p}} - \frac{\partial^2 \Gamma}{\partial \mathbf{p} \partial z}. \tag{64}$$

To eliminate the cross-derivative term, let us consider the vector derivative with respect to the transverse momentum of both sides of Eq. (60), i.e.

$$2\mathbf{p} + 2\frac{\partial \Gamma}{\partial z}\frac{\partial^2 \Gamma}{\partial \mathbf{p} \partial z} = -\left(\frac{\partial n^2}{\partial \mathbf{x}} \cdot \frac{\partial}{\partial \mathbf{p}}\right)\frac{\partial \Gamma}{\partial \mathbf{p}}. \tag{65}$$

The evaluation of this expression at $\mathbf{p} = \mathbf{P}(z, \xi)$ gives, after some reordering,

$$\frac{\mathbf{P}}{H} = -\frac{1}{2H}\left(\frac{\partial n^2}{\partial \mathbf{x}} \cdot \frac{\partial}{\partial \mathbf{p}}\right)\frac{\partial \Gamma}{\partial \mathbf{p}} - \frac{\partial^2 \Gamma}{\partial \mathbf{p} \partial z}. \tag{66}$$

From the comparison of Eqs. (64) and (66), we can see that it is convenient to choose

$$\dot{\mathbf{X}} = \frac{\mathbf{P}}{H}, \tag{67}$$

$$\dot{\mathbf{P}} = \frac{1}{2H}\frac{\partial n^2}{\partial \mathbf{x}}(\mathbf{X}, z). \tag{68}$$

These equations are identical to Eqs. (20) and (22). Also, notice that the substitution of Eqs. (62) and (63) into Eq. (60) evaluated at $\mathbf{p} = \mathbf{P}(z, \xi)$ implies that

$$H = \sqrt{n^2(\mathbf{X}, z) - \mathbf{P} \cdot \mathbf{P}}, \tag{69}$$

so the rays we obtain from the asymptotic treatment in the momentum representation are identical to those found in the position representation. This is a rather remarkable fact. It suggests why the ray model is so robust: it is obtained from wave optics in several limiting situations.

The phase function is again obtained parametrically. Let us define

$$T(z, \boldsymbol{\xi}) := \Gamma(\mathbf{P}, t). \tag{70}$$

The evolution of this function results from considering its derivative with respect to z:

$$\dot{T} = \frac{\partial \Gamma}{\partial \mathbf{p}} \cdot \dot{\mathbf{P}} + \frac{\partial \Gamma}{\partial z} = -\frac{1}{2H}\mathbf{X} \cdot \frac{\partial n^2}{\partial \mathbf{x}}(\mathbf{X}, z) + H. \tag{71}$$

The relation between the values of T for contiguous rays is found similarly by taking the partial derivative with respect to ξ_j of both sides of Eq. (70):

$$\frac{\partial T}{\partial \xi_j} = \frac{\partial \Gamma}{\partial \mathbf{p}} \cdot \frac{\partial \mathbf{P}}{\partial \xi_j} = -\mathbf{X} \cdot \frac{\partial \mathbf{P}}{\partial \xi_j}. \tag{72}$$

Now to solve the momentum transport equation in Eq. (61), let us define the differential operator in the mixed space $(\mathbf{p}, z)$ as

$$\tilde{\nabla} := \left(\frac{\partial}{\partial \mathbf{p}}, \frac{\partial}{\partial z}\right). \tag{73}$$

Let us also define the vector

$$\vec{\Theta} := (H\dot{\mathbf{P}}, H) = H\frac{\partial}{\partial z}(\mathbf{P}, z). \tag{74}$$

With this, Eq. (61) can be written as

$$\tilde{\nabla} \cdot (B_0^2 \vec{\Theta}) = 0. \tag{75}$$

This equation can be solved by integrating it over the volume in the $(\mathbf{p}, z)$ space occupied by an infinitesimal bundle of rays between z_0 and z, and using Gauss's theorem. Equation (74) states that $\vec{\Theta}$ is locally parallel to the rays in this space, so the contributions to the surface integral from the sides of the bundle vanish, as in the position-representation case. By following steps analogous to those in Subsection II.A.3, we find

$$\mathcal{B}_0(z, \boldsymbol{\xi}) := B_0(\mathbf{P}, z) = \sqrt{\frac{H(z_0, \boldsymbol{\xi})}{H(z, \boldsymbol{\xi})} \frac{\delta(\mathbf{P})}{\delta(\boldsymbol{\xi})}\bigg|_{z_0} \bigg/ \frac{\delta(\mathbf{P})}{\delta(\boldsymbol{\xi})}\bigg|_{z}}\; B_0(\mathbf{P}, z_0). \tag{76}$$

Notice that this solution has problems when the Jacobian in the denominator inside the square root vanishes. This happens when contiguous rays in the family have the same transverse momentum, i.e. when they are locally parallel. That is, the field estimate we are about to obtain also has problems, but these are different from those for the estimate found in the position representation, which happen at caustics. The location of these new problems, i.e. the places where contiguous rays have the same momentum, are called momentum caustics. As in the case of the amplitude of the position representation estimate, one must be very careful when choosing the sign of the square root in Eq. (76).

The field estimate is obtained by approximating $B \sim B_0$, i.e. $\tilde{U}(\mathbf{P}, z) \sim \mathcal{B}_0 \exp(ikT)$. To obtain the field in the position representation, we must calculate the inverse Fourier transform. Because we parametrized $\mathbf{p}$, the Fourier transform integral must be done parametrically, inserting a Jacobian factor:

$$U(\vec{r}) \sim \frac{k}{2\pi} \iint_{-\infty}^{\infty} B_0 \exp(ikT) \exp(ik\mathbf{x} \cdot \mathbf{P}) \frac{\delta(\mathbf{P})}{\delta(\xi)} d\xi_1 d\xi_2. \tag{77}$$

Notice that the contribution for each ray now extends over all the configuration space. That is, in this picture, rays do not contribute to the field as infinitesimally thin conduits of power. Instead, they are infinitely extended waves that interfere to make up the wave field. In fact, for the case of homogeneous media, these waves are plane waves. This estimate has no problems at caustics (focal points), but fails at momentum caustics, where the field (or a component of the field) is locally collimated.

IV. PHASE SPACE

We have learned about asymptotic approximations in both position and momentum representations. Surprisingly, both approaches lead to the same rays, although not to the same field estimates. It turns out that many other limits can be considered where, again, one finds the same rays but different transport-like equations. We will discuss some of these later.

At this point, it is convenient to introduce the concept of phase space. (Perhaps you are already familiar with this concept from other lectures in this school, and you will surely be exposed to it during the Winter College.) For simplicity, let us consider the case of two-dimensions. That is, we only have one transverse coordinate x. In this case, the ray equations become

$$\dot{X}(z,\xi) = \frac{P(z,\xi)}{H(z,\xi)}, \tag{78}$$

$$\dot{P}(z,\xi) = \frac{1}{2H(z,\xi)}\frac{\partial n^2}{\partial x}[X(z,\xi),z], \tag{79}$$

$$\dot{H}(z,\xi) = \frac{1}{2H(z,\xi)}\frac{\partial n^2}{\partial z}[X(z,\xi),z], \tag{80}$$

Notice that there is now only one parameter ξ that labels the rays. Additionally, we have the equations for the path lengths L and T, which become

$$\dot{L}(z,\xi) = \frac{n^2(X,z)}{H}, \tag{81}$$

$$L'(z,\xi) = PX', \tag{82}$$

$$\dot{T}(z,\xi) = H - \frac{X}{2H}\frac{\partial n^2}{\partial x}(X,z), \tag{83}$$

$$T'(z,\xi) = -XP', \tag{84}$$

where the primes denote derivatives in ξ.

At any fixed z, each ray is fully characterized by its transverse position X and its transverse momentum P: knowing where a ray is and where it is going is enough to trace it. We can then represent this ray by a point in a plane x vx p. This plane is called phase space. (Of course, for three dimensional fields, the phase space is four-dimensional, since there are two transverse directions and two transverse momenta. This is the reason why we start with the simpler two-dimensional case; the 2D phase space can be drawn on paper.) The complete ray family is then represented by a curve, traced by the points for each ray by varying ξ. This curve is called the phase space curve (PSC) or Lagrange manifold. Notice from Eq. (82) that the area under a segment of the phase space curve corresponds to the increment in the value of L between the rays that correspond to the ends of the PSC segment. This means that, if we know the phase space curve for a given z and the value of L for only one ray, we can determine the value of L for all the other rays. Similarly, it can be seen from Eq. (84) that the area to the left of a segment of the PSC (up to the p axis) corresponds to the difference in T for the corresponding end rays, so the knowledge of T for one ray and of the PSC is sufficient to find T for any other ray.

In the two-dimensional case, the formulas for the amplitudes of the estimates become

$$\mathcal{A}_0(z,\xi) = \sqrt{\frac{H(z_0,\xi)}{H(z,\xi)}\frac{X'(z_0,\xi)}{X'(z,\xi)}}A_0[X(z_0,\xi),z], \tag{85}$$

$$\mathcal{B}_0(z,\xi) = \sqrt{\frac{H(z_0,\xi)}{H(z,\xi)} \frac{P'(z_0,\xi)}{P'(z,\xi)}} B_0[P(z_0,\xi), z], \tag{86}$$

The field estimate resulting from using the position-dependent approach fails at caustics, i.e. when $\partial X/\partial\xi = 0$. In phase space, caustics correspond then to segments of the PSC that are locally vertical. On the other hand, the momentum-representation-based estimate fails at momentum caustics, when $\partial P/\partial\xi = 0$, i.e at segments of the PSC that are locally horizontal. One could formulate field estimates based on other representations associated, for example, with a fractional Fourier transform over the transverse variable of the field. These field estimates would be well behaved at both position and momentum caustics, but would fail around rays associated segments of the PSC with a given inclination (depending on the degree of the fractional Fourier transform). When a PSC is sufficiently complicated, the caustic problems are unavoidable, regardless of representation. Based on this fact, V.I. Maslov[8] proposed a scheme where the PSC is cut into pieces that are free of at least one type of caustic (position or momentum). A small transition region is left between the segments to avoid errors introduced by the abrupt cuts. A field estimate is then performed for each individual segment, using the appropriate prescription. The total estimate is found by adding the contributions due to all the segments. Notice that one must be careful to choose the correct Maslov index phase for each contribution. This approach is known as Maslov's canonical operator method, or simply Maslov's method[2]. Its implementation is, however, quite complicated, especially for three-dimensional fields, where instead of a PSC we have a two-dimensional Lagrange manifold embedded in a four-dimensional phase space. A different type of asymptotic method, which is free of problems at any type of caustics, and that is much easier to implement, will be discussed in the next section.

V. GAUSSIAN METHODS

For the position-based approach, the contribution associated with a ray is infinitesimally thin, while for the momentum-based approach, it is infinite in extent. Let us now formulate an estimate where each ray has a contribution that is finite, whose profile is given by some localized function G. For simplicity, let us start with the two-dimensional case. The field would then be written as a sum of contributions in the form

$$U(\vec{r}) = \int G[z,\xi,x - X(z,\xi)]d\xi, \tag{87}$$

where X is the transverse ray position. In what follows, we will use the names X, P, L, and H to refer to several functions that will appear in the derivations. While these functions will turn out to be the same ones found in previous sections, we will not assume so immediately.

To simplify things, let us assume a simple algebraic form for G. In particular, notice that it would be desirable for this contribution to look similar in both the position and momentum representations. One simple function that presents this property is a Gaussian:

$$G(z,\xi,x-X) = F(z,\xi)\exp\left[-\frac{K}{2}(x-X)^2\right]\exp[ikP(z,\xi)(x-X)], \tag{88}$$

where the linear phase factor was added so that the Fourier transform of G is a Gaussian in p centered at P. Notice that the width of the Gaussian in Eq. (88) is $K^{-1/2}$, while the width of its Fourier transform would be $K^{1/2}/k$. Because in the asymptotic limit we want both widths to behave similarly, we choose $K = k\gamma$, where γ is a parameter with units of inverse length. (We will discuss the effect of this parameter later.) This way, the width of the contribution is proportional to $1/\sqrt{k}$ for both the position and momentum representations. Also, let us assume that the prefactor independent of x, i.e. F, can be written in the now familiar way $F = w(z,\xi)\exp[ikL(z,\xi)]$, where L is real. The field construction in Eq. (87) can then be written as

$$U_\gamma(\vec{r}) = \int w(z,\xi)g(x,z,\xi)d\xi, \tag{89}$$

where

$$g(x,z,\xi) := \exp\left(-\frac{k\gamma}{2}[x-X(z,\xi)]^2 + ik\{L(z,\xi)+P(z,\xi)[x-X(z,\xi)]\}\right). \tag{90}$$

There are now two possibilities. One is to substitute each contribution in Eq. (89), i.e. wg, into the Helmholtz equation, and find the conditions for which this equation is satisfied, at least asymptotically. By doing this, it is found that X and P are indeed the position and momentum of a ray, and that L is its optical path length. (This will be shown later.) However, in order to force the remaining parts of the equation to vanish, one must assume that γ is a function of both z and ξ. The cancellation of the remaining leading orders leads to expressions for $\dot{w}$ and $\dot{\gamma}$. These expressions are nonlinear in these functions, and must in general be integrated numerically. Each contribution is then an independent "parabasal" (i.e. paraxial around a central ray) Gaussian beam centered at the ray corresponding to ξ. This type of method is known as a Gaussian beam summation (GBS), and several variants of it have been proposed independently within optics, quantum mechanics, and geology[9, 10].

The main problem with this approach is that, after some propagation, a beam's position and/or momentum can become very wide, violating the parabasal approximation.

The second approach, referred to as "stable aggregates of flexible elements" (SAFE)[11–15], consists on not separating the different contributions, but instead thinking of them as interrelated components. It turns out that, in this case, γ can be regarded as a constant (although one could also let it vary with z or ξ). To find this estimate, we substitute the full integral of contributions in Eq. (90) into Eq. (1). First, let us compute the derivative of U_γ with respect to x:

$$\frac{\partial U_\gamma}{\partial x} = \int w[-k\gamma(x-X) + ikP]g\,d\xi = \int wk(-\gamma\Delta + iP)g\,d\xi, \tag{91}$$

where $\Delta = x - X$. The second derivative in the same variable gives

$$\frac{\partial^2 U_\gamma}{\partial x^2} = \int w[-k\gamma + k^2(\gamma\Delta - iP)^2]g\,d\xi. \tag{92}$$

Similarly, the derivatives of U_γ with respect to z give

$$\frac{\partial U\gamma}{\partial z} = \int [\dot{w} + wk(\dot{Y}\Delta + iH)]g\,d\xi, \tag{93}$$

$$\frac{\partial^2 U\gamma}{\partial z^2} = \int \{\ddot{w} + 2\dot{w}k(\dot{Y}\Delta + iH) + w[k(\ddot{Y}\Delta - \dot{Y}\dot{X} + i\dot{H}) + k^2(\dot{Y}\Delta + iH)^2]\}g\,d\xi, \tag{94}$$

where $Y = \gamma X + iP$ and $H := \dot{L} - P\dot{X}$. Finally, it is convenient to expand the refractive index in a Taylor series around the ray's position as

$$n^2(x,z) = \sum_{j=0}^{\infty} N_j\Delta^j, \quad N_j := \frac{1}{j!}\frac{\partial^j n^2}{\partial x^j}(X,z). \tag{95}$$

We now substitute Eqs. (92), (94), and (95) into Eq. (1). The result can be written as

$$\int [k^2(C_{20} + C_{21}\Delta + C_{22}\Delta^2 + ...) + k(C_{10} + C_{11}\Delta) + C_{00}]\,g\,d\xi = 0, \tag{96}$$

where

$$C_{20} = w(N_0 - P^2 - H^2) = w[n^2(X,z) - P^2 - H^2], \tag{97}$$

$$C_{21} = w(N_1 - 2i\gamma P + 2iH\dot{Y}) = w\left[\frac{\partial n^2}{\partial x}(X,z) - 2H\dot{P} + 2i\gamma(H\dot{X} - P)\right], \tag{98}$$

$$C_{22} = w(N_2 + \gamma^2 + \dot{Y}^2), \tag{99}$$

$$C_{2j} = wN_j, \quad j \geq 3, \tag{100}$$

$$C_{10} = 2i\dot{w}H + w(-\gamma - \dot{Y}\dot{X} + i\dot{H}), \tag{101}$$

$$C_{11} = 2\dot{w}\dot{Y} + w\ddot{Y}, \tag{102}$$

$$C_{00} = \ddot{w}. \tag{103}$$

We now assume that k is large, and go on to perform the asymptotic treatment. Notice, however, that within the coefficients of the powers of k, the terms involving different powers of Δ cannot in principle be mixed, because Δ depends on the spatial variable x. Also, since g is a Gaussian of width $1/\sqrt{k\gamma}$ centered at $x = X$, and $\Delta = x - X$, the importance of the contributions to the integral due to each term of the integrand in Eq. (96) decreases with increasing j. The main contribution then is that corresponding to C_{20}. This contribution vanishes if we choose X, P, and H to be the ray quantities defined in the previous sections. If we do this, it turns out that C_{21} vanishes also, due to Eqs. (78) and (86). Again, making the leading orders of the asymptotic form of the wave equation vanish leads to the usual rays!

The integral in Eq. (96) can now be written as

$$\int [k^2(C_{22}\Delta^2 + ...) + k(C_{10} + C_{11}\Delta) + C_{00}]\, g\, d\xi = 0, \tag{104}$$

As an aside, it must be noted that the only difference up to now between SAFE and the GBS approach is the presence of the integral, together with the fact that the surviving coefficients in the case of GBS include extra terms involving derivatives of γ with respect to z. In particular, C_{22} includes the extra term $-i\dot{\gamma}H$. The basis of GBS is to eliminate the leading remaining contributions by forcing C_{22} and C_{10} to vanish. This leads, respectively, to the evolution equations for the width factor γ and the amplitude w. On the other hand, SAFE takes advantage of the presence of the integral, so that the two terms mentioned earlier can be mixed and give a simple result. This is done by using a trick based on the fact that the derivative of g with respect to ξ is given by

$$g' = k[\gamma X'\Delta + i(L' - PX' + P'\Delta)]g = k[Y'\Delta + i(L' - PX')]g. \tag{105}$$

Let us choose L to be the usual optical path length measured from a wavefront. Then, because of Eq. (82), $L' - PX' = 0$, and Eq. (105) can be written as

$$\Delta g = \frac{g'}{kY'}. \tag{106}$$

This expression, combined with integration by parts, can be used to remove the factors of Δ from Eq. (104). For each term one can do integration by parts in the form:

$$\begin{aligned}\int k^m C_{mj}\Delta^j g\, d\xi &= \int k^{m-1}\frac{C_{mj}}{Y'}\Delta^{j-1} g'\, d\xi = -\int k^{m-1}\left(\frac{C_{mj}}{Y'}\Delta^{j-1}\right)' g\, d\xi \\ &= \int k^{m-1}\left[C_{mj}\frac{X'}{Y'}\Delta^{j-2} - \left(\frac{C_{mj}}{Y'}\right)'\Delta^{j-1}\right] g\, d\xi,\end{aligned} \tag{107}$$

where the integrated terms in the integration by parts were dropped by assuming that the magnitude of the contributions go to zero in the limits. By using this trick repeatedly, we can rewrite Eq. (104) as

$$\int \left[k\left(\frac{X'}{Y'}C_{22} + C_{10}\right) + \mathcal{O}(k^0)\right] g\, d\xi = 0. \tag{108}$$

The leading term of this equation vanishes if $X'C_{22} + Y'C_{10} = \mathcal{O}(k^{-1})$. Let us now find out how to choose w for making this happen:

$$\begin{aligned} X'C_{22} + Y'C_{10} &= wX'(N_2 + \gamma^2 + \dot{Y}^2) + Y'[2i\dot{w}H + w(-\gamma - \dot{Y}\dot{X} + i\dot{H})] \\ &= w(N_1'/2 + \gamma^2 X' + \dot{Y}^2 X') + w(-\gamma Y' - \dot{Y}\dot{X}Y') + iY'(2\dot{w}H + w\dot{H}) \\ &= w[(H\dot{P})' + \gamma^2 X' + \dot{Y}^2 X' - \gamma Y' - \dot{Y}\dot{X}Y'] + iY'(2\dot{w}H + w\dot{H}) \\ &= w[(H\dot{P})' - i\gamma P' + \dot{Y}(\dot{Y}X' - \dot{X}Y')] + iY'(2\dot{w}H + w\dot{H}) \\ &= w[(H\dot{P} - i\gamma P)' + i\dot{Y}H'] + iY'(2\dot{w}H + w\dot{H}) \\ &= iw[\dot{Y}H' - (\gamma H\dot{X} + i2H\dot{P})'] + iY'(2\dot{w}H + w\dot{H}) \\ &= iw[\dot{Y}H' - (H\dot{Y})'] + iY'(2\dot{w}H + w\dot{H}) \\ &= -iwH\dot{Y}' + iY'(2\dot{w}H + w\dot{H}) \\ &= 2i\sqrt{HY'^3}\frac{\partial}{\partial z}\left(w\sqrt{\frac{H}{Y'}}\right), \end{aligned} \tag{109}$$

where we used the fact that $N_2 X' = N_1'/2$ and

$$H' = (\sqrt{n^2(X,z) - P^2})' = (N_1 X' - 2PP')/2H = \dot{P}X' - P'\dot{X} = -i(\dot{Y}X' - P'\dot{Y}). \tag{110}$$

The result of Eq. (109) is asympotically negligible if w is given by a Debye series of the form

$$w = \sqrt{\frac{Y'(z,\xi)}{H(z,\xi)}} \sum_{j=0}^{\infty} \frac{a_j}{(ik)^j}, \tag{111}$$

where the dominant term of the sum, i.e. $a_0(\xi)$ is independent of both γ and z.

The field can be estimated by approximating $a \sim a_0$, giving

$$U_\gamma(\vec{r}) \sim \int a_0 \sqrt{\frac{\gamma X' + iP'}{H}} \exp\left[-\frac{k\gamma}{2}(x - X)^2\right] \exp\{ik[L + (x - X)P]\}\, d\xi. \tag{112}$$

This estimate does not fail at caustics of any kind. Its only problem happens when rays turn around in z, i.e. if H vanishes. This is the basic result of SAFE. It would seem that this estimate depends strongly on the choice of γ. However, it is easy to show that this is not the case. To see this, consider the derivative of Eq. (112) with respect to γ, that is

$$\frac{\partial U_\gamma}{\partial \gamma} = \int a_0 \sqrt{\frac{Y'}{H}} \left(\frac{X'}{2Y'} - k\frac{\Delta^2}{2} \right) g \, d\xi = \int a_0 \sqrt{\frac{Y'}{H}} \left[\frac{X'}{2Y'} - k\frac{X'}{2Y'} + \mathcal{O}(k^{-1}) \right] g \, d\xi = \mathcal{O}(k^{-1}) U_\gamma, \tag{113}$$

where we used the integration by parts trick in the second step. This means that the variation of SAFE's estimate due to changes of the contributions' widths is asymptotically small. This is the reason for the name of the method. It is worth noticing that, in the limit when γ goes to zero, this estimate reduces to the momentum representation estimate, as the contributions become infinitely wide. On the other hand, when γ goes to infinity, the Gaussian contributions become delta functions, and the estimate reduces to that resulting from the position representation treatment presented in Section II. For finite, non-zero γ, we have an estimate that does not fail at any type of caustics.

To conclude, the generalization of this result to three dimensions is trivial. The field estimate is now given by

$$U_\gamma(\vec{r}) \sim \int a_0(\xi) \sqrt{\frac{1}{H} \frac{\delta(\gamma \cdot \mathbf{X} + i\mathbf{P})}{\delta(\xi)}} \, \exp\left[-\frac{k}{2} (\mathbf{x} - \mathbf{X}) \cdot \gamma \cdot (\mathbf{x} - \mathbf{X}) \right] \, \exp\{ik[L + (\mathbf{x} - \mathbf{X}) \cdot \mathbf{P}]\} \, d^2\xi. \tag{114}$$

Notice that, now, γ is in the most general case a 2×2 matrix.

APPENDIX A: MATHEMATICAL BACKGROUND: THE METHOD OF STATIONARY PHASE

Consider a general integral of the form

$$I(K) = \int_{-\infty}^{\infty} A(x) \, \exp[iK\Phi(x)] \, dx, \tag{A1}$$

where A and Φ are analytic, Φ is real, and K is a large parameter. Notice that, for sufficiently large K, the variation of the integrand is dominated by the exponential, which oscillates rapidly in x. These fast oscillations cause massive cancellation in the integral, except in the regions where the expression does not oscillate, that is, where $\Phi(x)$ is roughly constant. These regions are the vicinities of the so-called stationary points x_j, which satisfy

$$\Phi'(x_j) = 0. \tag{A2}$$

The main contributions to the integral in Eq. (A1) are then due to the vicinities of the stationary points.

Let us assume first, for simplicity, that there is only one stationary point x_0. A simple estimation of the integral follows from Taylor expanding A and Φ in Eq. (A1) around this stationary point:

$$\begin{aligned} I(K) &= \int_{-\infty}^{\infty} \left[A(x_0) + (x - x_0)A'(x_0) + \frac{(x-x_0)^2}{2} A''(x_0) + ... \right] \\ &\times \exp\left\{ iK \left[\Phi(x_0) + \frac{(x-x_0)^2}{2} \Phi''(x_0) + ... \right] \right\} dx \\ &= \exp[iK\Phi(x_0)] \int_{-\infty}^{\infty} \left[A(x_0) + \tau A'(x_0) + \frac{\tau^2}{2} A''(x_0) + ... \right] \\ &\times \exp\left\{ iK \left[\frac{\tau^2}{2} \Phi''(x_0) + ... \right] \right\} d\tau, \end{aligned} \tag{A3}$$

where we made the change of variables $\tau = x - x_0$. The main component of the result is found by truncating the series at the constant term for A and the quadratic term for Φ. The resulting integral can be solved in closed form giving

$$\begin{aligned} I(K) &\sim A(x_0) \exp[iK\Phi(x_0)] \int_{-\infty}^{\infty} \exp\left[\frac{iK\Phi''(x_0)\,\tau^2}{2} \right] d\tau \\ &= \sqrt{\frac{2\pi i}{K\Phi''(x_0)}}\, A(x_0) \exp[iK\Phi(x_0)] \\ &= \sqrt{\frac{2\pi}{K|\Phi''(x_0)|}}\, A(x_0) \exp\left\{ iK\Phi(x_0) + i\frac{\pi}{4}\mathrm{sgn}[\Phi''(x_0)] \right\}, \end{aligned} \tag{A4}$$

where sgn(a)= $a/|a|$. One can show that all the terms in both series that were neglected give only corrections to this result, which are smaller than the leading term by a factor proportional to a negative integer power of K. Therefore, for sufficiently large K, the estimate in Eq. (A4) is accurate.

When there are multiple stationary points, the estimate is given by the sum of the contributions of each of them:

$$I(K) \sim \sum_j \sqrt{\frac{2\pi}{K|\Phi''(x_j)|}}\, A(x_j) \exp\left\{ iK\Phi(x_j) + i\frac{\pi}{4}\mathrm{sgn}[\Phi''(x_j)] \right\}. \tag{A5}$$

The condition for using this formula is that the stationary points are mutually sufficiently separated, so that the integrand oscillates many times between them. The method of stationary phase can also be applied to multivariable integrals, but this falls beyond the scope of this course.

---

[1] M. Born and E. Wolf, *Principles of Optics*, 7th ed. (Cambridge University Press, Cambridge, 1999).

[2] Yu. A. Kravtsov and Yu. I. Orlov, *Caustics, Catastrophes, and Wave Fields*, 2nd ed., (Springer-Verlag, Heidelberg, 1998).

[3] A list of references is given in: M.C. Gutzwiller, "Resource Letter ICQM-1: The Interplay between Classical and Quantum Mechanics," *Am. J. Phys.* **66**, 304-324 (1998).

[4] See, for example, D. Bohm, *Quantum Theory*, (Dover, Mineola, 1989), pp. 264-295.

[5] J.H. Van Vleck, "The correspondence principle in the statistical interpretation of quantum mechanics," *Proc. N.A.S.* **14**, 178-188 (1928).

[6] M.C. Gutzwiller, *Chaos in Classical and Quantum Mechanics*, (Springer-Verlag, New York, 1990).

[7] D. Bohm, "A Suggested Interpretation of the Quantum Theory in Terms of Hidden Variables, I and II," *Phys. Rev.* **85**, 166-193 (1952).

[8] V.P. Maslov, *Perturbation Theory and Asymptotic Methods*, (Moskov UP, Moscow 1965).

[9] M.M. Popov, "A new method of computation of wave fields using Gaussian beams," *Wave Motion* **4**, 85-97 (1982).

[10] E.J. Heller, "Time-dependent approach to semiclassical dynamics," *J. Chem. Phys.* **62**, 1544-1555 (1975).

[11] G. W. Forbes and M. A. Alonso, "Using rays better. I. Theory for smoothly varying media," *J. Opt. Soc. Am. A* **18**, 1132-1145 (2001).

[12] M.A. Alonso and G.W. Forbes, "Using rays better. II. Ray families to match prescribed wave fields," *J. Opt. Soc. Am. A* **18**, 1146-1159 (2001).

[13] M. A. Alonso and G. W. Forbes, "Using rays better. III. Error estimates and illustrative applications in smooth media," *J. Opt. Soc. Am. A* **18**, 1357-1370 (2001).

[14] G.W. Forbes, "Using rays better. IV. Theory for refraction and reflection," *J. Opt. Soc. Am. A* **18**, 2557-2564 (2001).

[15] M. A. Alonso and G.W. Forbes, "Stable aggregates of flexible elements give a stronger link between rays and waves," *Opt. Exp.* **10**, 728-739 (2002).