Geometric Optics and Introduction to Fourier Integral Operators

Plamen Stefanov

Department of Mathematics, Purdue University, West Lafayette, IN 47907

November 8, 2026

TABLE OF CONTENTS

CHAPTER I. Introduction 1

CHAPTER II. The wave and the Helmholtz equations in open sets 5

§1. The initial value problem for the wave equation 5

§2. The geometric optics ansatz 7

§3. Solving the eikonal equation 9

§4. Solving the transport equations 18

§5. The parametrix and wave front sets 19

§6. Justification of the parametrix 20

§7. High frequency waves 20

§8. Two parameter expansions; beams 23

§9. The Schwartz kernel of the solution operator of the wave equation. A generalized Kirchhoff formula 23

§10. "Half-wave equation". General hyperbolic ΨDO equations of real principal type . . . 25

§11. Helmholtz equations 27

CHAPTER III. Introduction to Fourier Integral Operators through geometric optics 31

§1. The wave front set of the local solution 31

§2. The canonical relation of the wave propagator 34

§3. Other phase functions 40

§4. Fourier integral operators 41

§5. Changing phase functions and caustics 43

§6. Continuation through a spherical focus 45

§7. The propagator on the sphere 47

§8. Global operators and continuation of the propagator 52

CHAPTER IV. Gaussian beams 55
§1. The Euclidean case 55
§2. The metric case. 63
§3. The wave equation 68
CHAPTER V. The wave equation in a domain with boundary 71
§1. The 1D case 71
§2. The wave equation with a boundary source. Incoming and outgoing solutions 73
§3. The Cauchy problem on the lateral boundary 76
§4. Reflection of waves 78
§5. The transmission problem 79
CHAPTER VI. The Helmholtz equation in a domain with boundary 85
Roadmap 85
CHAPTER VII. The elasticity system 87
CHAPTER VIII. Hyperbolic systems 89
CHAPTER IX. Weakly non-linear geometric optics 91
Bibliography 93

Chapter I

Introduction

The purpose of these notes is to present the geometric optics construction for wave equations and derive some of its consequences. We discuss geometric optics for the Helmholtz equation in the high-frequency regime as well. The methods extend to other hyperbolic equations, for example to the isotropic elasticity system. These notes are based on a course I gave at the Math Department at Purdue University. The first version was written well before AI was a thing but they have been polished by AI now. Some new additions are AI generated and clearly marked as such. I plan to rewrite them eventually.

We start with the acoustic wave equation

$$(\partial_t^2 - c^2(x)\Delta)u = 0, \quad (t,x) \in \mathbf{R} \times \mathbf{R}^n, \tag{0.1}$$

where $c(x) > 0$ is a smooth wave speed. Let us look at several ways of constructing solutions.

Parametrix for the Cauchy problem. Suppose we are given u and u_t at $t = 0$. We look for a solution built out of terms of the form

$$u(t,x) \sim (2\pi)^{-n} \int e^{\mathrm{i}\phi(t,x,\xi)} a(t,x,\xi) \hat{f}(\xi)\, \mathrm{d}\xi, \tag{0.2}$$

where f comes from the Cauchy data, a is a classical amplitude, and the phase function ϕ is positively homogeneous of degree one in ξ. This is our first example of a *Fourier integral operator*, or FIO. We will introduce its geometry by following the rays determined by the phase. Plug (0.2) into (0.1) and equate the leading homogeneous terms to get the eikonal equation

$$\phi_t^2 = c^2 |\nabla_x \phi|^2. \tag{0.3}$$

The successive lower order terms give us transport equations for the expansion of a. If we take $\phi|_{t=0} = x \cdot \xi$ and $a|_{t=0} = 1$, then (0.2) at $t = 0$ is just the Fourier inversion formula for f. There are two choices $\phi_\pm$ for ϕ in (0.3) coming from the two square roots in (0.3), which tells us that we should seek a solution as a sum of two terms as (0.2). The initial conditions for u and u_t allow us to choose the amplitudes of the two terms to match the two Cauchy data. Details are given in the next chapter.

The amplitude in expansion (0.2) is determined up to $a \in S^{-\infty}$. It is easy to see, similarly to the ΨDO calculus, that this means a smooth error. Therefore, the r.h.s. of (0.2) has the same singularities as the exact solution.

If $c = 1$, the two phases are $\phi_\pm = x \cdot \xi \mp t|\xi|$, see (II.1.10). A singularity starting at y with covector ξ travels along $x = y \pm t\xi/|\xi|$ on the corresponding branch. For variable c, the spatial projections of the null bicharacteristics are geodesics of the metric $c^{-2}\mathrm{d}x^2$. The transport equations are ODEs along the bicharacteristics. We can therefore follow a singularity and compute its leading amplitude along the ray.

Progressive wave expansion. Such expansions appear in scattering theory. Assume that $c(x) = 1$ for $|x| \geq R$ for some $R > 0$. Fix $\omega \in S^{n-1}$. At time t, the incident plane wave $\delta(t - x \cdot \omega)$ is supported on the plane $x \cdot \omega = t$, moving in the direction ω. For $t < -R$ the plane has yet to reach $\mathrm{supp}(c - 1)$. We look for the solution of

$$(\partial_t^2 - c^2(x)\Delta)u = 0, \quad u(t, x) = \delta(t - x \cdot \omega) \quad \text{for } t < -R. \tag{0.4}$$

Near the incoming front we look for an expansion of the form

$$u \sim \sum_{j=0}^{\infty} a_j(x, \omega) h_j(t - \psi(x, \omega)), \tag{0.5}$$

where $h_0 = \delta$, h_1 is the Heaviside function, and $h_{j+1}(s) = s_+^j/j!$ for $j \geq 1$. Plug (0.5) into (0.4) and collect terms of the same order of singularity. The phase $t - \psi(x, \omega)$ must solve (0.3) again, which gives

$$c^2|\nabla_x \psi|^2 = 1, \quad \psi|_{x\cdot\omega=-R} = x \cdot \omega. \tag{0.6}$$

There are two choices for the normal derivative on this plane. Matching ψ to the incident phase $x \cdot \omega$ on the incoming side gives $\partial_\omega \psi = 1$ there. The amplitudes solve transport equations. The r.h.s. of (0.5) represents the exact solution up to a smooth error.

High-frequency stationary waves. In the Cauchy parametrix we describe singularities up to a smooth remainder. Now we consider a family of waves depending on a large frequency λ and measure the error by its decay as $\lambda \to +\infty$. Each wave can be smooth for fixed λ and the family can still have a semiclassical wave front set.

Take the formal Fourier transform of (0.1) w.r.t. t. We get the acoustic Helmholtz equation

$$(-c^2\Delta - \lambda^2)\hat{u}(\lambda, x) = 0 \quad \text{in } \mathbf{R}^n. \tag{0.7}$$

We look for solutions of the form

$$\hat{u}(\lambda, x) = e^{\mathrm{i}\lambda\phi(x)} a(x, \lambda), \tag{0.8}$$

where a is a classical amplitude in the semiclassical sense,

$$a(x, \lambda) \sim \sum_{k=0}^{\infty} \lambda^{-k} a_k(x). \tag{0.9}$$

See also (II.7.1) for the time-dependent case. Plug (0.8) into (0.7). The coefficient of λ^2 gives $c^2|\nabla_x \phi|^2 = 1$, just as in (0.6). The choice of phase depends on the boundary data or, in a scattering problem, on the incident wave.

High-frequency time-dependent waves. In Chapter II, Section 7, we fix a direction ω and study individual oscillatory waves as $\lambda \to +\infty$. We seek approximate solutions of (0.1) of the form

$$u(t, x) = e^{\mathrm{i}\lambda\phi(t,x,\omega)} a(t, x, \omega, \lambda), \tag{0.10}$$

where ϕ is real and $a \sim \sum_{k=0}^{\infty} \lambda^{-k} a_k(t,x,\omega)$, as in (II.7.1). Substitution gives the eikonal equation (0.3) at order λ^2 and successive transport equations for the a_k along the rays.

For $c = 1$, take $\phi = t - x \cdot \omega$, $|\omega| = 1$, and prescribe $a_0(0,x) = \chi(x)$ with $\chi \in C_0^\infty$ supported near x_0. Then $a_0(t,x) = \chi(x - t\omega)$, so the leading wave packet moves along $x = x_0 + t\omega$. The lower order amplitudes correct the error caused by transverse localization. A single branch also determines the initial velocity; to match arbitrary Cauchy data we use both branches. For variable c the same construction follows the geodesics of $c^{-2}\mathrm{d}x^2$ while the phase remains regular. On a fixed time interval within such a patch, keeping N terms gives an $O(\lambda^{1-N})$ error in (u, u_t) in $H^1 \times L^2$, provided both Cauchy data of the exact solution and the approximation are matched.

Gaussian beams. How closely can we concentrate a high-frequency wave near a single ray? In Chapter II, Section 8, we introduce a second parameter $\varepsilon > 0$ by replacing the initial cutoff by $\chi((x - x_0)/\varepsilon)$. At order k, the transport construction involves up to $2k$ derivatives of the cutoff, so the effective expansion parameter is $(\lambda\varepsilon^2)^{-1}$. Thus we need $\lambda\varepsilon^2 \to \infty$; for example, $\varepsilon = \lambda^{-\alpha}$ with $0 < \alpha < 1/2$ works. At the critical scale $\varepsilon = \lambda^{-1/2}$, these estimates no longer give decreasing orders.

For the wave equation, one way to see this scale is to start with initial data of coherent state type,

$$f_\lambda(x) = e^{\mathrm{i}\lambda(x-x_0)\cdot\omega - \lambda|x-x_0|^2/2}. \tag{0.11}$$

The Gaussian has "width" of order $\lambda^{-1/2}$ about x_0, and the oscillatory factor gives the packet its central wave vector $\lambda\omega$. We include the Gaussian in a complex phase with a positive imaginary quadratic part. To construct a Gaussian beam, we solve the eikonal and transport equations to the required order along the ray. The imaginary part gives decay away from it. If we follow one half-wave, the center travels along the corresponding ray.

Chapter IV develops this construction and also gives a stationary example for the Euclidean Helmholtz equation in $\mathbf{R}^3$. A source on the plane $x^3 = 0$ has a Gaussian profile in the transverse variables x'. We apply the outgoing Green's function to this source and use complex stationary phase to compute the leading field on the beam axis. This connects the Gaussian beam construction with the outgoing Helmholtz waves discussed above.

In the next chapter we develop geometric optics for the wave and Helmholtz equations in open sets. Chapter III introduces Fourier integral operators, and Chapter IV treats Gaussian beams. We then ask what happens when a ray meets a boundary or an interface. This leads us to reflected and transmitted waves.

[The introduction is incomplete]

Chapter II

The wave and the Helmholtz equations in open sets

1. The initial value problem for the wave equation

We are interested in the wave equation

$$(\partial_t^2 - c^2\Delta_{g_0})u = 0 \tag{1.1}$$

in $\mathbf{R}_t \times \mathbf{R}_x^n$, where g_0 is a Riemannian metric and $c > 0$ is a smooth wave speed. For g_0 Euclidean this is the acoustic wave equation of Chapter I. The Cauchy problem with data at $t = 0$ is to solve (1.1) with the initial conditions

$$u|_{t=0} = f_1, \quad u_t|_{t=0} = f_2. \tag{1.2}$$

We are interested in boundary value problems as well, and transmission problems, when the coefficients can jump over smooth hypersurfaces.

The principal symbol of (1.1) is

$$p(t, x, \tau, \xi) = -\tau^2 + c^2(x)|\xi|_{g_0}^2,$$

where $|\xi|_{g_0}^2 = g_0^{ij}(x)\xi_i\xi_j$. It is convenient to introduce the metric $g = c^{-2}g_0$. Then $|\xi|_g = c|\xi|_{g_0}$, and away from the zero section the characteristic variety has two sheets $\Sigma_\pm = \{\tau \pm |\xi|_g = 0\}$. Let us see first how these two sheets enter the solution of the Cauchy problem.

1.1. The constant coefficients case. Assume g_0 Euclidean and c constant.

The 1D case. Let $n = 1$ first. Consider the constant coefficients wave equation

$$u_{tt} - c^2u_{xx} = 0. \tag{1.3}$$

Since $\partial_t^2 - c^2\partial_x^2 = (\partial_t - c\partial_x)(\partial_t + c\partial_x)$, one easily gets that

$$u = F(x - ct) + G(x + ct) \tag{1.4}$$

with F and G arbitrary (distributions) represents the general solution. We have two waves moving with speeds c (and velocities c and $-c$, respectively), see Figure 1. Having initial conditions, one can easily compute them. If $f_2 = 0$, for example, then $F = G = \frac{1}{2}f_1$. If f_1 is singular, we see that

each singularity splits into two equal parts and then each one of them propagates to the left and right with speed c. The solution cannot be a ΨDO (applied to f_1). Even though this is a very basic

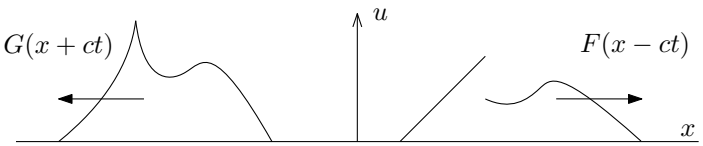


Figure 1. propagation of 1D waves.

case ($n = 1$, constant coefficients), this example describes the behavior of the multidimensional variable coefficient case as we show later, see Figure 6.

The multidimensional case. Let $n \geq 2$. Make the speed c equal to 1. One has explicit formulas then, of course but let us see what happens if we use Fourier analysis. We write

$$(\partial_t^2 + |\xi|^2)\hat{u}(t,\xi) = 0,$$

where $\hat{u}$ is the partial Fourier transform w.r.t. x. This is an ODE, therefore,

$$\hat{u}(t,\xi) = a_1(\xi)e^{\mathrm{i}t|\xi|} + a_2(\xi)e^{-\mathrm{i}t|\xi|}. \tag{1.5}$$

Using the initial conditions, we get

$$\hat{u}(t,\xi) = \frac{1}{2}e^{\mathrm{i}t|\xi|}\left(\hat{f}_1(\xi) + \frac{\hat{f}_2(\xi)}{\mathrm{i}|\xi|}\right) + \frac{1}{2}e^{-\mathrm{i}t|\xi|}\left(\hat{f}_1(\xi) - \frac{\hat{f}_2(\xi)}{\mathrm{i}|\xi|}\right). \tag{1.6}$$

Differentiating at $t = 0$ gives $\hat{f}_2$, as it should. Combining the two terms, we get

$$\hat{u}(t,\xi) = \cos(t|\xi|)\hat{f}_1(\xi) + \frac{\sin(t|\xi|)}{|\xi|}\hat{f}_2(\xi). \tag{1.7}$$

At $\xi = 0$ the sine quotient has the limiting value t. Taking the inverse Fourier transform, we get

$$\begin{aligned} u(t,x) = (2\pi)^{-n}\int e^{\mathrm{i}(-t|\xi|+x\cdot\xi)}\left(\frac{1}{2}\hat{f}_1(\xi) - \frac{1}{2\mathrm{i}|\xi|}\hat{f}_2(\xi)\right)\mathrm{d}\xi \\ + (2\pi)^{-n}\int e^{\mathrm{i}(t|\xi|+x\cdot\xi)}\left(\frac{1}{2}\hat{f}_1(\xi) + \frac{1}{2\mathrm{i}|\xi|}\hat{f}_2(\xi)\right)\mathrm{d}\xi. \end{aligned} \tag{1.8}$$

This can also be written as

$$\begin{aligned} u(t,x) = (2\pi)^{-n}\iint e^{\mathrm{i}(-t|\xi|+(x-y)\cdot\xi)}\left(\frac{1}{2}f_1(y) - \frac{1}{2\mathrm{i}|\xi|}f_2(y)\right)\mathrm{d}y\,\mathrm{d}\xi \\ + (2\pi)^{-n}\iint e^{\mathrm{i}(t|\xi|+(x-y)\cdot\xi)}\left(\frac{1}{2}f_1(y) + \frac{1}{2\mathrm{i}|\xi|}f_2(y)\right)\mathrm{d}y\,\mathrm{d}\xi. \end{aligned} \tag{1.9}$$

We get oscillatory integrals. They are not ΨDOs! The essential difference with the ΨDOs is the phase functions above: we have

$$\phi_\pm := \mp t|\xi| + x\cdot\xi, \tag{1.10}$$

respectively $\mp t|\xi| + (x-y)\cdot\xi$ instead of $x\cdot\xi$ or $(x-y)\cdot\xi$.

Of course, we also have the classical formulas for the solutions using spherical means, etc.

We can already see how singularities propagate. The condition for the phase in (1.9) to be stationary in ξ is $x = y \pm t\xi/|\xi|$. Away from these rays we can integrate by parts in ξ. To follow a singularity at y in a given direction, localize near y and to large frequencies in a small conic neighborhood of that direction. On each branch the Fourier multiplier $e^{\mp \mathrm{i}t|\xi|}$ preserves the frequency localization. Thus the covector stays fixed and the base point moves along $x = y \pm t\xi/|\xi|$, just as in Chapter I.

1.2. The variable coefficients case. Let us return to (1.1). The spatial operator $P := -c^2\Delta_{g_0}$ is formally self-adjoint for the measure $\mathrm{d}\mu = c^{-2}\mathrm{d}\,\mathrm{Vol}_{g_0}$, where $\mathrm{d}\,\mathrm{Vol}_{g_0} = \sqrt{\det g_0}\,\mathrm{d}x$. For the global statements on $\mathbf{R}^n$ below, assume for simplicity that g_0 is Euclidean and $c = 1$ outside a compact set. Given a domain U, and a function $u(t,x)$, define the energy

$$E_U(t,u) = \int_U \left(|\nabla u|^2_{g_0} + c^{-2}|u_t|^2\right) \mathrm{d}\,\mathrm{Vol}_{g_0},$$

where $|\nabla u|^2_{g_0} = g_0^{ij}(\partial_i u)\overline{\partial_j u}$. By finite speed of propagation, compactly supported data give a solution with compact spatial support on each bounded time interval. The domain of dependence is determined by $g = c^{-2}g_0$. The energy norm for the Cauchy data (f_1, f_2), that we denote by $\|\cdot\|_{\mathcal{H}}$ is then defined by

$$\|(f_1,f_2)\|^2_{\mathcal{H}} = \int_{\mathbf{R}^n} \left(|\nabla f_1|^2_{g_0} + c^{-2}|f_2|^2\right) \mathrm{d}\,\mathrm{Vol}_{g_0},$$

This defines the energy space

$$\mathcal{H} = H_D(\mathbf{R}^n) \oplus L^2(\mathbf{R}^n, \mathrm{d}\mu),$$

where $H_D(\mathbf{R}^n)$ is the completion of $C_0^\infty(\mathbf{R}^n)$ in the norm $\left(\int |\nabla f|^2_{g_0}\,\mathrm{d}\,\mathrm{Vol}_{g_0}\right)^{1/2}$. The wave equation then can be written down as the system

$$\mathbf{u}_t = \mathbf{P}\mathbf{u}, \quad \mathbf{P} = \begin{pmatrix} 0 & I \\ -P & 0 \end{pmatrix}, \tag{1.11}$$

where $\mathbf{u} = (u, u_t)$. Take the nonnegative self-adjoint realization of P defined by the Dirichlet energy form in $L^2(\mathrm{d}\mu)$. On a bounded smooth domain the same construction with form domain H_0^1 gives the homogeneous Dirichlet boundary condition. The corresponding wave generator is skew-adjoint on the energy space. Stone's theorem then gives

$$\mathbf{u}(t,\cdot) = e^{t\mathbf{P}}\mathbf{f}, \tag{1.12}$$

where $\mathbf{f} = (f_1, f_2)$. The group $e^{t\mathbf{P}}$ is unitary (energy preserving) and strongly continuous.

We now know that the Cauchy problem is solvable and that the energy is preserved. Let us see where a singularity travels and how its leading amplitude changes.

2. The geometric optics ansatz

Before we construct the phases of Chapter I, let us allow lower order terms as well. For a smooth one-form A set

$$\Delta_{g_0,A} = (\det g_0)^{-1/2}(\partial_{x^i} - \mathrm{i}A_i)(\det g_0)^{1/2} g_0^{ij}(\partial_{x^j} - \mathrm{i}A_j). \tag{2.1}$$

We consider the equation

$$(\partial_t^2 + a\partial_t - c^2\Delta_{g_0,A} + q)u = 0 \tag{2.2}$$

with Cauchy data (1.2). Here a is the attenuation and q is a potential. The coefficients of the lower order terms may be complex. Another example with the same principal part for g_0 Euclidean is the acoustic pressure equation

$$p_{tt} - c^2\rho\nabla\cdot(\rho^{-1}\nabla p) = 0, \tag{2.3}$$

where $\rho > 0$ is the density. Expanding the divergence gives a first order term involving $\nabla \log \rho$. In all these examples the principal symbol, and hence the rays, are the same for fixed c and g_0. The lower order terms enter the amplitude equations. It is convenient first to write the t derivatives in divergence form as well. Set

$$A_0 = \frac{1}{2}a.$$

It is straightforward to show that

$$\partial_t^2 + a\partial_t = (\partial_t + A_0)^2 - \partial_t A_0 - A_0^2.$$

Therefore, setting

$$V := -\partial_t A_0 - A_0^2 + q,$$

we can rewrite (2.2) as

$$Pu = 0, \quad P := (\partial_t + A_0)^2 - c^2\Delta_{g_0,A} + V. \tag{2.4}$$

From here on P denotes this spacetime operator.

Motivated by the constant coefficient case, we look for a solution of (2.4), (1.2) of the type (1.8). This leads to the ansatz

$$\begin{aligned} u(t,x) = (2\pi)^{-n}\sum_{\sigma=\pm}\int e^{\mathrm{i}\phi_\sigma(t,x,\xi)}\Big(&a_{1,\sigma}(t,x,\xi)\hat f_1(\xi) \\ &+ |\xi|_g^{-1}a_{2,\sigma}(t,x,\xi)\hat f_2(\xi)\Big)\mathrm{d}\xi, \end{aligned} \tag{2.5}$$

modulo terms involving smoothing operators of f_1 and f_2. Here the phase functions $\phi_\pm$ are positively homogeneous of order 1 in ξ. Next, we seek the amplitudes in the form

$$a_{j,\sigma} \sim \sum_{k=0}^\infty a_{j,\sigma}^{(k)}, \quad \sigma = \pm,\ j = 1,2, \tag{2.6}$$

where $a_{j,\sigma}^{(k)}$ is homogeneous in ξ of degree $-k$ for large $|\xi|$. To construct such a solution, we plug (2.5) into (2.4) and try to kill all terms in the expansion in homogeneous (in ξ) terms. The following identity, which we can obtain by a direct calculation, helps with that.

Lemma 2.1.

$$\begin{aligned} e^{-i\phi}Pe^{i\phi} = &-(\partial_t\phi)^2 + c^2 g_0^{jk}(\partial_j\phi)(\partial_k\phi) \\ &+ i(\partial_t^2 - c^2\Delta_{g_0})\phi + 2i\Big((\partial_t\phi)(\partial_t + A_0) - c^2 g_0^{jk}\partial_j\phi(\partial_k - \mathrm{i}A_k)\Big) + P, \end{aligned} \tag{2.7}$$

where $e^{\pm i\phi}$ and all other stand-alone functions stand for operators of multiplication.

Proof. It is easy to see that it is enough to prove (2.7) for $V = 0$. We have

$$(\partial_k - \mathrm{i}A_k)\, e^{\mathrm{i}\phi}u = e^{\mathrm{i}\phi}\left((\partial_k - \mathrm{i}A_k) + \mathrm{i}\partial_k\phi\right)u. \tag{2.8}$$

Apply (2.8) twice, keeping the density $(\det g_0)^{1/2}$ inside the outer derivative. The terms in which that derivative falls on $\partial_k\phi$ or on the metric density combine to give the ordinary Laplacian of ϕ. We get

$$e^{-\mathrm{i}\phi}\Delta_{g_0,A}e^{\mathrm{i}\phi}u = \Delta_{g_0,A}u + \mathrm{i}(\Delta_{g_0}\phi)u + 2\mathrm{i}g_0^{jk}(\partial_j\phi)(\partial_k - \mathrm{i}A_k)u - |\nabla_x\phi|^2_{g_0}u. \tag{2.9}$$

Similarly,

$$e^{-\mathrm{i}\phi}(\partial_t + A_0)^2e^{\mathrm{i}\phi}u = (\partial_t + A_0)^2u + \mathrm{i}(\partial_t^2\phi)u + 2\mathrm{i}(\partial_t\phi)(\partial_t + A_0)u - (\partial_t\phi)^2u. \tag{2.10}$$

Multiply (2.9) by $-c^2$ and add it to (2.10) to complete the proof. □

Equating the terms of order 2 yields the *eikonal equation*

$$(\partial_t\phi)^2 - c^2(x)|\nabla_x\phi|^2_{g_0} = 0. \tag{2.11}$$

Here $|\nabla_x\phi|^2_{g_0} = g_0^{ij}\partial_i\phi\partial_j\phi$ is the squared norm of $\nabla_x\phi$ as a covector. Clearly, $c^2|\nabla_x\phi|^2_{g_0} = |\nabla_x\phi|^2_g$.

Write $f_j = (2\pi)^{-n}\int e^{\mathrm{i}x\cdot\xi}\hat{f}_j(\xi)\,\mathrm{d}\xi$, $j = 1, 2$, to get the following initial conditions for $\phi_\pm$

$$\phi_\pm|_{t=0} = x\cdot\xi. \tag{2.12}$$

For the transport equations we must use the full amplitude multiplying each datum in (2.5). Write $b^{(k)}_{1,\pm} = a^{(k)}_{1,\pm}$ and $b^{(k)}_{2,\pm} = |\xi|_g^{-1}a^{(k)}_{2,\pm}$. The factor $|\xi|_g^{-1}$ depends on x for a variable metric. Equating the terms one order below the eikonal equation gives the first *transport equation*

$$\Big((\partial_t\phi_\pm)(\partial_t + A_0) - c^2g_0^{ij}(\partial_i\phi_\pm)(\partial_j - \mathrm{i}A_j) + C_\pm\Big)\,b^{(0)}_{j,\pm} = 0, \tag{2.13}$$

with

$$2C_\pm = (\partial_t^2 - c^2\Delta_{g_0})\phi_\pm. \tag{2.14}$$

At each lower order we get the same transport operator, now with a right-hand side computed from the preceding coefficient. We will return to these equations after solving for the phase.

We will solve the eikonal equation first. Note that if we have the general differential operator with a principal symbol $p(x,\xi)$, the eikonal equation would be

$$p(x, \nabla\phi) = 0.$$

This is also known as (one of the forms of) the Hamilton-Jacobi equation.

3. Solving the eikonal equation

3.1. The Hamilton-Jacobi equation: the method of characteristics. Let $H(x,\xi)$ be a smooth Hamiltonian (so far just a smooth function) of $(x,\xi) \in \mathbf{R}^n \times \mathbf{R}^n$. We want to solve the Hamilton-Jacobi equation (eikonal) with prescribed values on some hypersurface S

$$H(x, \nabla\phi) = 0, \quad \phi|_S = \phi_0. \tag{3.1}$$

We work locally near some $x_0 \in S$.

There are several versions of that equation which are equivalent to (3.1). The most common form is

$$\partial_t\phi + G(x, \nabla_x\phi) = 0. \tag{3.2}$$

This form appears in geometric optics constructions for hyperbolic PDEs. Add the variable $x^0 = t$ and its dual variable τ. Setting $H = \tau + G(x, \xi)$ gives (3.1) in spacetime. For Cauchy data on $t = 0$ the characteristic field is transverse to the initial surface since $H_\tau = 1$.

Another version is

$$G(x, \nabla\phi) = 1. \tag{3.3}$$

This is the eikonal equation for Hemholtz type of PDEs with a large (or small) parameter. For example, $g^{ij}(x)\phi_{x^i}\phi_{x^j} = 1$, with g a Riemannian metric. Write $H = G - 1$ to get (3.1).

To solve (3.1), we use the method of characteristics, as in [4]. It is described there how to solve the more general equation $H(x, \nabla\phi, \phi) = 0$. In our case, there are some minor simplifications.

The idea is to find some curves (characteristics) $x(s)$ so that we can easy compute ϕ restricted to those characteristics: $\phi(x(s))$. Those curves would hit S, where we have an initial value ϕ_0. They would depend on an additional $n - 1$ dimensional parameter, for example on the point of intersection with S.

The form of (3.1) suggests that those curves might be the Hamiltonian curves for H, i.e., the solutions of

$$\dot{x} = H_\xi, \quad \dot{\xi} = -H_x. \tag{3.4}$$

Then (3.1) would just reflect the fact that H is constant along those curves if we can arrange

$$\xi(t) = (\nabla\phi)(x(t)) \tag{3.5}$$

(meaning $\nabla_x\phi$ evaluated at $x = x(t)$). This is just a wild guess, and let us see if we can make it happen. Assume for a moment that (3.1) is locally solvable. Let $x(s)$ be some smooth curve and take (3.5) as a definition of ξ without assuming that it has anything to do with (3.4). Differentiate w.r.t. t to get

$$\dot{\xi}_i(t) = \phi_{x^i x^j}(x(t))\dot{x}^j. \tag{3.6}$$

We have unpleasant second derivatives here and to deal with them, differentiate (3.1) to get

$$H_{x^i} + H_{\xi_j}\phi_{x^j x^i} = 0, \quad \forall i. \tag{3.7}$$

Comparing the last two equations, we see that it is very convenient to take $\dot{x} = H_\xi$ to make the expressions involving the second derivatives the same. Once we have done that, the last equation becomes $\dot{\xi}(t) = -H_x(x(t), \xi(t))$. Those are exactly the Hamiltonian equations (3.4).

How to find ϕ? Set $z(t) = \phi(x(t))$. Then

$$\dot{z}(t) = \nabla\phi \cdot \dot{x}(t) = \xi(t) \cdot H_\xi(x(t), \xi(t)). \tag{3.8}$$

To start this ODE we need the initial covector on S. Its tangential part is $\mathrm{d}\phi_0$. Choose a conormal ν and any extension η of $\mathrm{d}\phi_0$ to the ambient cotangent space at x_0. We solve $H(x_0, \eta + \alpha\nu) = 0$ for α. At the chosen root assume $H_\xi \cdot \nu \neq 0$. The implicit function theorem then gives a smooth choice of root nearby, and $\dot{x} = H_\xi$ is transverse to S. This is the noncharacteristic condition for this branch. If there are two simple roots, each gives a local solution.

Let us carry out the construction. Solve

$$\dot{z} = \xi \cdot H_\xi \tag{3.9}$$

along the Hamiltonian curves with initial conditions $z(0, x_0) = \phi_0(x_0)$ on $t = 0$ defining S locally: The map $(t, x_0) \mapsto x$ is a local diffeomorphism, and we set $\phi(x) = z(t, x_0)$. We will prove (3.5), i.e.,

that $\nabla\phi(x(t,x_0)) = \xi(t,x_0)$ which is true on $t=0$ by construction. Differentiate (3.9) with respect to x_0. We get

$$\phi_{x^i}\partial_{x_0^j}x^i(t,x_0) = \partial_{x_0^j}\phi_0 + \int_0^t \Big(\partial_{x_0^j}\xi(s,x_0)\cdot\dot{x}(s,x_0) + \xi(s,x_0)\cdot\partial_{x_0^j}\dot{x}(s,x_0)\Big)\,\mathrm{d}s, \tag{3.10}$$

$j=1,\dots,n-1$. Integrate by parts the second term. The new integrand is

$$\partial_{x_0^j}\xi(s,x_0)\cdot\dot{x}(s,x_0) - \dot{\xi}(s,x_0)\cdot\partial_{x_0^j}x(s,x_0) = \partial_{x_0^j}H(x(s,x_0),\xi(s,x_0)) = 0.$$

The boundary terms are

$$\xi(t,x_0)\cdot\partial_{x_0^j}x(t,x_0) - \xi(0,x_0)\cdot\partial_{x_0^j}x(0,x_0).$$

The last term above is the tangential projection of $-\xi(0,x_0)$ onto $T_{x_0}S$. It cancels the $\partial_{x_0^j}\phi_0$ term in (3.10). Accounting for all those cancellations, we get

$$(\phi_{x^i}(x(t,x_0)) - \xi_i(t,x_0))\partial_{x_0^j}x^i(t,x_0) = 0, \quad j=1,\dots,n-1.$$

This implies that the projection of $\nabla\phi - \xi$ to the hyperplane spanned by $v_j := \partial_{x_0^j}x^i(t,x_0)$, $j=1,\dots,n-1$, vanishes. By (3.9), the projection to $H_\xi = \dot{x}$ vanishes as well, and that vector is transversal to the previous plane for $|t| \ll 1$. Therefore, (3.5) holds.

Finally, we show that ϕ does satisfy (3.1). We know that $H(x(t),\xi(t)) = 0$ for every bicharacteristic as long as this is true for $t=0$. By (3.5), equation (3.1) holds along each such bicharacteristics.

Set $\phi(x(t)) = z(t)$. Define $\xi(t)$ by (3.5). Differentiate to get (3.6) which we can also write as $-H_{x^i} = \dot{\xi}_i = \phi_{x^ix^j}H_{\xi_j}$. This implies

$$\frac{\mathrm{d}}{\mathrm{d}t}H(x(t),\nabla\phi(x(t))) = 0.$$

If we can arrange (3.1) on S, then we have it along $x(t)$ as well.

We emphasize again that this solution is valid as long as $(x',s)\mapsto x$ is a diffeomorphism. In general, that is true in some neighborhood of x_0 only.

Finally, if H is homogeneous of order m in ξ, then $\xi\cdot H_\xi = mH$, which vanishes on the zero energy level.

This gives us the following recipe for solving (3.1) locally near x_0, see also Figure 2.

- Find $\nabla\phi$ on S when S is non-characteristic (for the given ϕ_0). Fix a locally smooth solution, if more than one. We get a covector field $(x',\xi(x'))$, with $x'\in S$ and $\xi(x') = \nabla\phi(x')$.
- Solve the Hamiltonian system (3.4) with initial conditions given by the covector field above. That defines the bicharacteristics $(x(s),\xi(s))$. They lie on the zero energy level $H=0$.
- Solve $\dot{z} = \xi\cdot H_\xi$ along the characteristics $x = x(s)$ with initial condition $z(0) = \phi(x',0)$ in the coordinates (x',s). Then $\phi = z$ in those coordinates.
- Pass to the original coordinates x to get ϕ.

We will use (3.5) repeatedly since it identifies the transported covector with the differential of the phase.

We consider some special cases now.

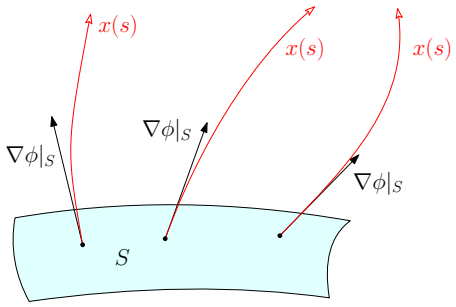


Figure 2. Solving the Hamilton-Jacobi equation. We solve $\dot{z} = \xi \cdot H_\xi$ along each characteristic.

3.2. Example: the Helmholtz equation. If we are looking for a solution of the Helmholtz equation

$$(-\Delta_g - \lambda^2)u = 0$$

of the form $u = e^{\mathrm{i}\lambda\phi(x)}a(x,\lambda)$ with $a(x,\lambda)$ having an asymptotic expansion in λ, one gets the following eikonal equation

$$g^{ij}\phi_{x^i}\phi_{x^j} = 1, \quad \phi|_S = \phi_0. \tag{3.11}$$

If we have $c^2\Delta$ instead, we have to replace g by $c^{-2}g$. The eikonal equation can be written as $|\nabla\phi|_g^2 = 1$, where $|\cdot|_g$ is the norm of the covector $\mathrm{d}\phi$ at x. The Hamiltonian then is $H(x,\xi) = \frac{1}{2}\left(g^{ij}(x)\xi_i\xi_j - 1\right)$. The factor $1/2$ is just for convenience. We can also say that the Hamiltonian is $G(x,\xi) = \frac{1}{2}g^{ij}(x)\xi_i\xi_j$ and we work on the energy level $G = 1/2$. Then (3.9) takes the form $\dot{z} = 1$ because $\xi \cdot H_\xi = g^{ij}\xi_i\xi_j = 1$. Then along the characteristics, $\phi = \phi_0 + s$.

The Hamiltonian system take the form

$$\dot{x}^i = g^{ij}\xi_j, \quad \dot{\xi}_i = -\frac{1}{2}(\partial_{x^i}g^{kl})\xi_k\xi_l. \tag{3.12}$$

Equation (3.11) implies that those curves lie on the zero energy level of H, i.e., they are zero (null) bicharacteristics of H. The projections $x(s) = \gamma(s)$ are unit speed geodesics in g, and $\xi_i(s) = g_{ij}\dot{\gamma}^j(s)$ are the corresponding covector components.

So the solution can be described like this. Solve the eikonal equation at S first, which is equivalent to finding the normal derivative $\nabla_\nu\phi$ on S. It is known that there exist local semigeodesic coordinates in which $S = x^n = 0$, x^n is the signed distance to S (in the metric) in a choosen orientation, and the metric takes the form $g_{\alpha\beta}\mathrm{d}x^\alpha\mathrm{d}x^\beta + (\mathrm{d}x^n)^2$, where α, β take values from 1 to $n-1$, and ∂_{x^n}. The eikonal equation then takes the form

$$g^{\alpha\beta}\phi_{x^\alpha}\phi_{x^\beta} + \phi_{x^n}^2 = 1, \quad \phi|_{x^n=0} = \phi_0(x'). \tag{3.13}$$

On S, we have either two solutions for ϕ_{x^n}, or one, or zero, depending on the sign of $1 - g^{\alpha\beta}\phi_{x^\alpha}\phi_{x^\beta}|_S = 1 - |\nabla'\phi_0|_{g'}^2$, where ∇' is the tangential differential, and g' is the restriction of g on the boundary.

(i) The hyperbolic region: $|\nabla'\phi_0|_{g'}^2 < 1$ Assume this inequality near some point. Then on S, (3.13) has two solutions:

$$\phi_{x^n}\big|_{x^n=0} = \pm\sqrt{1 - g^{\alpha\beta}(\phi_0)_{x^\alpha}(\phi_0)_{x^\beta}}\big|_{x^n=0}.$$

Choose one of those solutions in a smooth way near x_0, for example by requiring $\phi_{x^n} > 0$. Then $\nabla\phi$ determines the covector field $\xi = \nabla\phi|_S$, see (3.5) with base points on S. Note that this covector field is unit (in the metric $\{g^{ij}\}$ on $T^*\mathbf{R}^n$) because the eikonal equation says just that. For the incident front of Chapter I, $S = \{x \cdot \omega = -R\}$ and the tangential differential vanishes. With ω as the normal, matching the incident phase gives the positive root. We now issue bicharacteristics from the chosen covectors. In the (x', s) coordinates, $\phi = \phi_0(x') + s$. Physically, this corresponds to

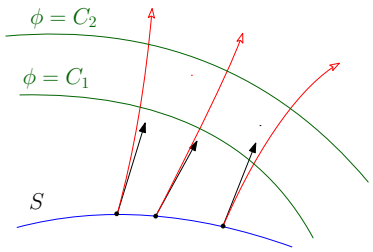


Figure 3. Solving $|\nabla\phi|_g^2 = 1$. The characteristics are unit speed geodesics. The solution increases as the arc length (in the metric) along each characteristic. The level hypersurfaces $\{\phi = \text{const.}\}$ are perpendicular (in the metric) to the characteristics. Note that there is another local solution with characteristics going down, like a "distorted mirror image" of the top ones.

a wavefront moving away from S in the positive ($x^n > 0$) direction. We can call it outgoing. If we choose the other solution on S, with $\phi_{x^n} < 0$, we get a wave coming to S, we can call it incoming.

Equality (3.5) in this case has a nice geometric meaning. By the Hamiltonian equations (3.12), $\xi = g\dot{x}$ along the bicharacteristics. Then we get

$$\dot{x} = g^{-1}\nabla\phi \tag{3.14}$$

along the bicharacteristics. The expression on the right is the covector field $\nabla\phi$ identified with a vector one by the metric. In Riemannian geometry, this is called $\operatorname{grad}\phi$. We thus get $\dot{x} = \operatorname{grad}\phi$. Therefore, the characteristics $x(t)$ are integral curves of teh gradient field $\operatorname{grad}\phi$ and in particular, they are perpendicular (as always, in the metric) to the level hypersurfaces of ϕ.

Assume for a moment that g is Euclidean. We do not get an explicit solution for general S. When $S = \{x^n = 0\}$, and try to solve

$$|\nabla\phi|^2 = 1, \quad \phi|_{x^n=0} = x \cdot \xi',$$

we get explicit the solutions $\phi_\pm = x \cdot \xi_\pm$, where $\xi_\pm = (\xi', \pm\sqrt{1 - |\xi'|^2})$ if $|\xi'| < 1$.

As we saw above, $\phi = x \cdot \omega$ solves the eikonal equation (without worrying about boundary conditions) for every unit ω. The corresponding solutions $u = e^{\mathrm{i}\lambda x\cdot\omega}$ are called harmonic plane waves "propagating" in the direction ω. Another candidate for that would be $u = e^{-\mathrm{i}\lambda x\cdot\omega}$ and the choice of the sign is a matter of convention. In dimension three, another special solution is $\phi = |x|$ for $x \neq 0$. Then we get $u = e^{\mathrm{i}\lambda|x|}a$ and it turns out that $a = 1/|x|$ works as an amplitude, so $u = e^{\mathrm{i}\lambda|x|}/|x|$ is a solution for $x \neq 0$. Also, $u = e^{-\mathrm{i}\lambda|x|}/|x|$ is a solution. In the whole space, up to a constant, they are actually fundamental solutions of the Helmholtz operator. They are called spherical waves. They satisfy a constant boundary condition on any sphere centered at the origin.

(ii) The elliptic region: $|\nabla'\phi_0|^2_{g'} > 1$. In this case, (3.13) has no real solutions on S. It has two complex ones however:

$$\phi_{x^n}\big|_{x^n=0} = \pm \mathrm{i}\sqrt{g^{\alpha\beta}(\phi_0)_{x^\alpha}(\phi_0)_{x^\beta} - 1}\big|_{x^n=0}.$$

A solution away from S may not exist. The characteristics are complex now, and complex vector fields may lead to non-solvable linear PDEs. What is done is this case is to solve the eikonal equation up to an error $O((x^n)^\infty)$ near $x^n = 0$ for $x^n > 0$. The transport equations later would be solved up to such an error, as well. Then in the justification part of the parametrix, we need to use that fact, and this will be done later.

(iii) The glancing region: a neighborhood of $|\nabla'\phi_0|^2_{g'} = 1$. We really need to work in a neighborhood because the sign of $g^{\alpha\beta}(\phi_0)_{x^\alpha}(\phi_0)_{x^\beta} - 1$ may change between the zeros, and the zeros can have a singular behavior, for example, they may have an accumulation point. This case is much more delicate. Assuming strict convexity of S, one needs to change the type of the ansatz (2.5): instead of exponential functions, we use Airy ones. This construction is beyond the scope of those notes.

3.3. Example: the wave equation. The eikonal equation for the wave equation

$$(\partial_t^2 - \Delta_g)u = 0$$

is

$$(\partial_t\phi)^2 - g^{ij}\phi_{x^i}\phi_{x^j} = 0, \quad \phi|_S = \phi_0. \tag{3.15}$$

If g is euclidean, it is just $\phi_t^2 - |\nabla_x\phi|^2 = 0$. Even in the metric case, it still looks like that but $|\cdot|$ is the norm of a covector now.

If g is Euclidean, the phases $\phi_\pm = x\cdot\xi \mp t|\xi|$ give the exact solutions $e^{\mathrm{i}\lambda\phi_\pm}$. Write $\omega = \xi/|\xi|$ and absorb $|\xi|$ into the frequency. Superposing $e^{\mathrm{i}\lambda(x\cdot\omega\mp t)}$ over λ with measure $\mathrm{d}\lambda/(2\pi)$ gives the plane fronts $\delta(x\cdot\omega\mp t)$, moving in the directions $\pm\omega$.

In the general case, we apply the scheme above. Think of the t variable as another x one, and let τ be its dual one. The Hamiltonian is

$$H(t,x,\tau,\xi) = \frac{1}{2}\left(-\tau^2 + |\xi|_g^2\right). \tag{3.16}$$

The Hamiltonian system then takes the form

$$\dot t = -\tau, \quad \dot x = g^{-1}\xi, \quad \dot\tau = 0, \quad \dot\xi = -\nabla_x|\xi|_g^2/2. \tag{3.17}$$

This system can be decoupled. The second and the fourth equation coincide with (3.12); hence the (x,ξ) curves (no t and τ) are the bicharecteristics we had above and $x(s)$ are geodesics. Since τ is constant by the third equation, we also get $t = -s\tau$, where s is the parameter along the Hamiltonian curves. Since we work on the energy level $H = 0$, we have $|\tau| = |\xi|_g = \text{const}$. The speed $|\dot x|_g$ equals $|\tau| = \text{const}$. If we use t as a parameter along the bicharacteristics, we would get

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \pm\frac{g^{-1}\xi}{|\xi|_g}, \quad \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\xi_i}{|\xi|_g}\right) = \mp\frac{(\partial_{x^i}g^{jk})\xi_j\xi_k}{2|\xi|_g^2}, \tag{3.18}$$

where choice of the sign $\pm$ is the opposite of the sign of τ. Those curves are just the unit speed geodesics (after the identification of vectors and covectors) in the phase space.

To solve (3.15) at S, we first distinguish two conditions. Let $\nu = (\nu_t, \nu_x)$ be a conormal to S. The wave operator is non-characteristic at S if $\nu_t^2 \neq |\nu_x|_g^2$. If $\nu_t^2 > |\nu_x|_g^2$, the conormal is timelike

and S is spacelike, as for $S = \{t = 0\}$. If $\nu_t^2 < |\nu_x|_g^2$, the conormal is spacelike and S is timelike, as for a cylinder $\mathbf{R}_t \times S_0$.

What do we need for the eikonal equation? The covector $\zeta = (\tau, \xi) = \mathrm{d}\phi$ must be lightlike and have the prescribed tangential part on S. At the chosen root we need $\partial_\zeta H \cdot \nu \neq 0$, with H given by (3.16). As in the method of characteristics above, this says that the root is simple and its ray crosses S transversely. For example, at the flat wall $x^n = 0$ in Euclidean space the conormal is spacelike, but at a glancing covector $\xi_n = 0$ and the ray is tangent to the wall. To solve (3.15) at S, **assume first that S is spacelike**. The problem can then be described in the following way: we are given $\nabla' \phi_0$ (the tangential gradient on S) at a point on S near a fixed one. Is there a lightlike covector which projects to $\nabla' \phi_0$? For nonzero tangential data there are two such covectors.

For example, on $S = \{t = 0\}$ we have $\partial_t \phi_\pm = \mp |\nabla_x \phi_0|_g$. With the convention of Chapter I the plus branch has $\tau < 0$ and $\dot{t} > 0$ in (3.17). In flat geometry this gives (1.10). The general case can be reduced to this one by a suitable Lorentzian "rotation" around (t_0, x_0).

This gives us another point of view. The eikonal equation (3.15) is equivalent to the following two:

$$\partial_t \phi_\pm = \mp |\nabla_x \phi_\pm|_g. \tag{3.19}$$

Assume $\mathrm{d}\phi_0|_{TS} \neq 0$ for spacelike S. The two roots are then distinct. The corresponding Hamiltonians are $H_\pm = \tau \pm |\xi|_g$, with $H_+ H_- = -2H$. On either sheet the other factor is nonzero, so the Hamilton vector fields of H and of $H_\pm$ on $H_\pm = 0$ trace the same curves up to reparameterization.

We have $\nabla \phi$ on S now, and we can solve the eikonal equation. We have a homogeneous Hamiltonian now, and on the zero energy level, the right hand side of (3.9) vanishes! So ϕ is just constant along the characteristics (geodesics) $x(s)$ issued from points on S with initial directions $\dot{x} = g^{-1}\xi$ on S, where ξ is computed in the first step. Note that if $\gamma(s)$ is a geodesic in the x-space with speed $|\dot{\gamma}| = c$, we think of $(\pm ct, \gamma(t))$ as a geodesic in time-space.

The analog of (3.14) now is (by (3.5) and (3.17))

$$(\partial_t \phi, \nabla_x \phi) = (\tau, \xi) = (-\dot{t}, g\dot{x}) \tag{3.20}$$

along the bicharacteristics, with $\dot{t} = -\tau$ constant.

Example 3.1. If g is Euclidean, and $S = \{t = 0\}$, and $\phi|_S = x \cdot \xi$, we first determine $\partial_t \phi|_S = \pm|\xi|$, which gives us $\nabla_{t,x} \phi|_S = (\pm|\xi|, \xi)$. Then ϕ is constant along the lines $(t, x) = (|\xi| s, x_0 \pm s\xi)$ (with tangent vectors $(|\xi|, \pm\xi)$) with an initial condition $\phi_0 = x_0 \cdot \xi$ for $s = 0$. Since $s = t/|\xi|$, $x = x_0 \pm s\xi$, we get

$$\phi_\pm = (x \mp t\xi/|\xi|) \cdot \xi = \mp t|\xi| + x \cdot \xi.$$

Those are the phase functions we considered above.

Example 3.2. A more general and often needed case is to solve (3.15) with $S = \{t = 0\}$ and $\phi_0 = x \cdot \xi$, for $\xi \neq 0$ a parameter, i.e., solve

$$(\partial_t \phi)^2 - g^{ij} \phi_{x^i} \phi_{x^j} = 0, \quad \phi|_{t=0} = x \cdot \xi.$$

The choice $\phi_0 = x \cdot \xi$ is made in local coordinates. At $t = 0$ we have $\mathrm{d}_x \phi = \xi$ and $\partial_t \phi = \pm|\xi|_g$, where $|\xi|_g = \sqrt{g^{ij}(x)\xi_i \xi_j}$. Take the negative root. The bicharacteristic issued from $(0, x, -|\xi|_g, \xi)$ projects to $(s|\xi|_g, \gamma_{x,\xi}(s))$, since $\dot{t} = -\tau$ in (3.17). Along it the phase keeps its initial value $x \cdot \xi$. Thus $\phi(s|\xi|_g, \gamma_{x,\xi}(s), \xi) = x \cdot \xi$ or, equivalently,

$$\phi(t, \gamma_{x,\xi/|\xi|_g}(t), \xi) = x \cdot \xi \tag{3.21}$$

because $\gamma_{x,\xi}(t/|\xi|_g) = \gamma_{x,\xi/|\xi|_g}(t)$ by the homogeneity of the Hamiltonian (3.16) and as a result, the homogeneity of the Hamiltonian equations (3.17) w.r.t. ξ.

Compare this to the previous example, where g is Euclidean. Then the bicharacteritic in question is $s \mapsto (t = s|\xi|, x + s\xi, \tau = -|\xi|, \xi)$ and as mentioned above, we take $\gamma_{x,\xi}(s) = x + s\xi = x + t\xi/|\xi|$. Then (3.21) holds for $\phi(t, x, \xi) = -t|\xi| + x \cdot \xi$. This solution can be obtained directly from (3.21) by solving

$$\phi(t, x + t\xi/|\xi|, \xi) = x \cdot \xi$$

with the substitution $y = x + t\xi/|\xi|$. In the general case we fix t and ξ and invert the map $x \mapsto \gamma_{x,\xi/|\xi|_g}(t)$. At $t = 0$ this is the identity, so the implicit function theorem gives a local inverse for small $|t|$.

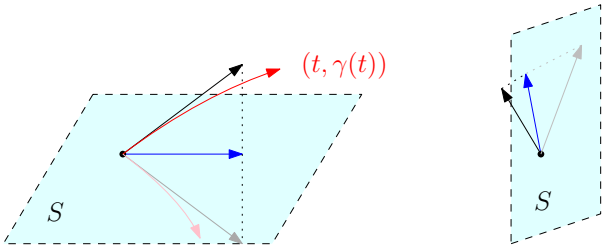


Figure 4. Solving $(\partial_t\phi)^2 - |\nabla\phi|_g^2 = 0$. The characteristics are geodesics $(t, \gamma(t))$ in time-space. The solution is constant along each characteristic; therefore equal to its value ϕ_0 on S. When S is spacelike (left), there are always two lighlike (co)vectors with a given projections; i.e., $\nabla\phi|_S$ has two locally smooth solutions. When S is timelike (right), there might be two, one or no such solution depending on whether $\partial'\phi|_S$ is timelike/ lightlike/ spacelike.

Assume next that S is timelike. If $S = \mathbf{R}_t \times S_0$ is a cylinder, for example, we are given $\nabla'\phi_0 = (\partial_t\phi_0, \nabla'_x\phi_0)$, where $\partial'_x\phi_0$ is the tangential derivative on S_0. A lightlike covector with such a projection may not exist however. Indeed, we need to determine $\partial_{x^n}\phi$ from the equation $(\partial_t\phi)^2 = (\nabla'_x\phi)_g^2 + (\partial_{x^n}\phi)^2$ in boundary normal coordinates but that is possible only if $(\partial_t\phi)^2 > (\nabla'_x\phi)_g^2$ (we want to avoid an equality). In other words, $\nabla'\phi_0|_S$ must be timelike (on S)! This is the analog of the hyperbolic region in the Helmholtz case. At glancing the normal component vanishes and the tangential covector itself is lightlike. The same description holds at a point of a general timelike hypersurface in an adapted Lorentzian frame.

After we are done with the boundary determination, we have a covector field $(x, \nabla\phi)$ on S. We solve (3.9) as in the spacelike case with initial conditions that field. In (3.9), we replace ξ by (τ, ξ). We have a homogeneous Hamiltonian and on the zero energy level, the right hand side of (3.9) vanishes as above. So ϕ is just constant along the characteristics (geodesics) $x(s)$ issued from points on S with initial directions $\dot{x} = g^{-1}\xi$ on S, where ξ is computed in the first step. As above, if $\gamma(s)$ is a geodesic in the x-space with speed $|\dot{\gamma}| = c$, we think of $(\pm ct, \gamma(t))$ as a geodesic in timespace.

A special type of solutions of (3.15) is

$$\phi(t, x) = t \mp \psi(x) \tag{3.22}$$

up to multiplication by a constant, where ψ solves (3.11). For the Euclidean initial phase $x\cdot\xi$, take $\psi = x\cdot\xi/|\xi|$. Then $|\xi|(\psi\mp t) = x\cdot\xi\mp t|\xi|$, as before. For a variable metric $|\mathrm{d}(x\cdot\xi)|_g$ generally varies with x, and this separation need not work. On the cylinder $S = \mathbf{R}_t\times S_0$, on the other hand, boundary data $t\mp\psi_0$ lead directly to (3.22) by solving the spatial eikonal equation with $\psi|_{S_0} = \psi_0$.

If $\nabla'\phi_0|_S$ is spacelike, a real phase function does not exist and even a complex valued one may not exist. Then one construct a solution up to an $O((x^n)^\infty)$ error, as in the Helmholtz case.

In the neighborhood of a lightlike $\nabla'\phi_0|_S$, a different construction is needed.

3.4. Example: the Lorentzian case. Let g be a Lorentzian metric in the "timespace" $\mathbf{R}^{1+n}$ with a signature $(-,+,\dots,+)$. An example is the Minkowski metric $= -\mathrm{d}t^2+(dx^1)^2+\cdots+(dx^n)^2$, where we think of t as x^0. The corresponding wave equation takes the form

$$|\det g|^{-1/2}\partial_{x^i}|\det g|^{1/2}g^{ij}\partial_{x^j}u = 0.$$

Of course, we can cancel the term on the left but written this way, the operator $\Box_g$ applied to u is an invariantly defined hyperbolic operator, formally looking as the Laplace-Beltrami operator in the Riemannian case. Looking for solutions of the kind (2.5), with (t,x) replaced by x there, we arrive at the eikonal equation

$$g^{ij}\phi_{x^i}\phi_{x^j} = 0,$$

which looks like (3.11) but not quite because there is zero on the right. It is actually a generalization of (3.15). Indeed, the wave equation in section 3.3 corresponds to the Lorentzian metric

$$-\mathrm{d}t^2 + g_{ij}(x)\mathrm{d}x^i\mathrm{d}x^j,$$

hence it is a special case of the more general Lorentzian case. Without going into details, let us just mention that the analysis is not so different than the one in section 3.3. In particular, the characteristics are null geodesics (satisfying $g_{ij}\dot\gamma^i\dot\gamma^j = 0$) and the phase is constant along them.

3.5. Caustics in the eikonal equation. As mentioned above, the eikonal equations are solvable in general locally only. We demonstrate below that indeed, the eikonal equation may not be solvable globally. Those effects are known as caustics.

Let us start with a simple example: solve $|\nabla\phi|^2 = 1$ with boundary data $\phi = 1$ on the unit sphere S. One solution is $\phi = |x|$ (another one is $\phi = 2-|x|$). We could have started with $\phi = C$ for any fixed constant C, including $C = 0$. The characteristics are easy to compute: they are the radial rays $r = s$ in polar coordinates. Reparameterize them as $r = 1-s$ so that they start from S and move to the origin. Then $|x| = 1-s$ would be the level surfaces of ϕ up to $s = 1$, corresponding to $x = 0$. We cannot extend the solution further in a smooth way. Even if S is a part of the unit sphere only; we would have the same problem. The wave fronts develop a caustic at the origin.

What happens to the rays at the origin? Their spatial projections meet, but they carry different covectors there. Each ray continues along its own straight line. It is the projection of this family of rays to x-space that becomes singular.

A boundary value problem for the wave equation related to this example would be to solve the constant coefficients wave equation $(\partial_t^2-\Delta)u = 0$ in the cylinder $\mathbf{R}\times B$, where B is the unit ball in $\mathbf{R}^3$ with boundary the unit sphere S as above. Send a spherical front inward from S at $t = 0$. Before it reaches the center, $u = \delta(t+|x|-1)/|x|$ solves the wave equation in the ball for $0<t<1$. Its front is $|x| = 1-t$, so it focuses at the origin at $t = 1$. The incoming expression alone cannot continue the homogeneous solution through the focus. In this example one can see the continuation

explicitly by subtracting $\delta(t-|x|-1)/|x|$: the point sources at the origin cancel, and the second front travels outward after $t=1$.

Other types of caustics are shown in Figure 5. The curves which are supposed to be the level surfaces of ϕ are either non-smooth (with a vertex) or they self-intersect or they may intersect other such surfaces (which is a contradiction). If a non-smooth curve (hypersurface) is a level set of ϕ, than at the non-smooth points, $\nabla\phi=0$. This is a contradiction with the eikonal equation.

Does focusing of this front mean that the Cauchy parametrix (2.5) also breaks down? In the Euclidean case we have the exact formula (1.8) with phases $x\cdot\xi\mp t|\xi|$ for every t, including the time of focusing. It is the radial phase $t+|x|-1$ of the incoming spherical front that ceases to be smooth at $x=0$. For a variable metric, the phase of the propagator can itself require a different representation when the map we inverted after (3.21) becomes singular. In either case we can still follow the rays in phase space.

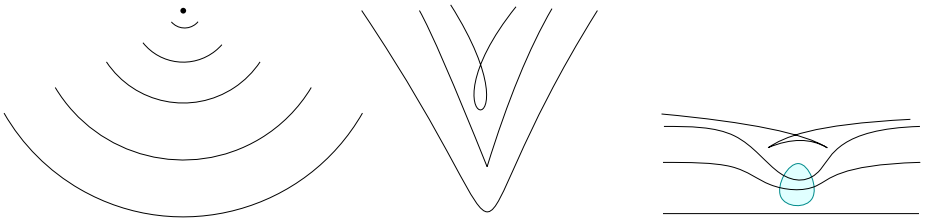


Figure 5. Caustics. Left: a wavefront focusing at a point; Euclidean geometry. Center: cusp that can form even if the geometry is Euclidean. Right: a wave moving upward with a slow region in the center.

4. Solving the transport equations

Let us return to (2.5), keeping the speed c explicit, so $g=c^{-2}g_0$. Denote

$$L_\pm := (\partial_t\phi_\pm)(\partial_t+A_0) - c^2 g_0^{ij}(\partial_i\phi_\pm)(\partial_j - \mathrm{i}A_j) + C_\pm, \tag{4.1}$$

$2C_\pm = (\partial_t^2 - c^2\Delta_{g_0})\phi_\pm$. Apply Lemma 2.1 to the full amplitudes introduced in (2.13). Once the phase solves (2.11), the conjugated operator is $2\mathrm{i}L_\pm + P$. Thus

$$L_\pm b_{j,\pm}^{(0)} = 0, \quad j=1,2. \tag{4.2}$$

For the next terms, we get

$$L_\pm b_{j,\pm}^{(k)} = -\frac{1}{2\mathrm{i}} P b_{j,\pm}^{(k-1)}, \quad k\geq 1, \quad j=1,2, \tag{4.3}$$

and the expressions on the right have been computed in the previous steps.

The vector field in $L_\pm$ is $(\partial_t\phi_\pm, -g^{-1}\nabla_x\phi_\pm)$. By (3.20) it is the negative of $(\dot t,\dot x)$ along the projected bicharacteristic $\Gamma(s)=(t(s),x(s))$ of (3.17). Thus (4.2) and (4.3) are linear ODEs on the very same rays. For example, if $A_0=0$ and $b=b_{j,\pm}^{(0)}$, the first equation reads

$$\frac{\mathrm{d}}{\mathrm{d}s} b(\Gamma(s)) = \big(C_\pm(\Gamma(s)) + \mathrm{i}A_k(\Gamma(s))\dot x^k(s)\big) b(\Gamma(s)).$$

Integrating gives the leading amplitude in terms of its value at $s = 0$. We can use t as a parameter as well since $\dot{t} = -\tau \neq 0$.

The initial conditions in (1.1) provide initial conditions for the transport equations. We get

$$a_{1,+} + a_{1,-} = 1, \quad a_{2,+} + a_{2,-} = 0 \quad \text{for } t = 0.$$

This is true in particular for the leading terms $a^{(0)}_{1,\pm}$ and $a^{(0)}_{2,\pm}$. Since $\partial_t \phi_\pm = \mp|\xi|_g$ at $t = 0$, differentiating the exponentials in (2.5) and matching $u_t = f_2$ at leading order gives

$$a^{(0)}_{1,+} = a^{(0)}_{1,-}, \quad \mathrm{i}(a^{(0)}_{2,-} - a^{(0)}_{2,+}) = 1 \quad \text{for } t = 0.$$

Therefore,

$$a^{(0)}_{1,+} = a^{(0)}_{1,-} = \frac{1}{2}, \quad a^{(0)}_{2,+} = -a^{(0)}_{2,-} = \frac{\mathrm{i}}{2} \quad \text{at } t = 0. \tag{4.4}$$

For the Euclidean wave operator with $c = 1$, this recovers (1.8), including the factor $|\xi|^{-1}$ multiplying $\hat{f}_2$. In general these initial values determine the leading $b^{(0)}_{j,\pm}$. At each lower order we again match both Cauchy data, now including the known time derivatives of the preceding amplitudes. We then integrate (4.3) and sum the resulting symbol expansions asymptotically.

5. The parametrix and wave front sets

Let us sketch how the parametrix compares singularities at $t = 0$ and at another time. We keep t fixed and take one branch of (2.5), of the form

$$\begin{aligned} u_0(t,x) &= (2\pi)^{-n} \int e^{\mathrm{i}\phi(t,x,\xi)} a(t,x,\xi) \hat{f}(\xi)\, \mathrm{d}\xi \\ &= (2\pi)^{-n} \iint e^{\mathrm{i}(\phi(t,x,\xi) - y\cdot\xi)} a(t,x,\xi) f(y)\, \mathrm{d}\xi\, \mathrm{d}y. \end{aligned} \tag{5.1}$$

Here $f \in \mathcal{E}'$, the amplitude is a symbol, and we work on a small conic patch with low frequencies removed. Recall that $(y,\xi) \notin \mathrm{WF}(f)$ means that, after localization near y, the Fourier transform of f decreases faster than any power in a cone about ξ.

Write $A_\pm(t)f = u_0(t,\cdot)$ for the two choices of phase. To test the output near x in the covector direction η, localize there and take its Fourier transform. The phase becomes $\phi_\pm(t,x,\xi) - y\cdot\xi - x\cdot\eta$. Integration by parts gives rapid decrease unless

$$y = \partial_\xi \phi_\pm(t,x,\xi), \qquad \eta = \partial_x \phi_\pm(t,x,\xi).$$

Only input directions in $\mathrm{WF}(f)$ can contribute: in the other directions the localized Fourier transform of the datum already decreases rapidly. Thus the possible output singularities are obtained by following these stationary equations.

Put $b(x,\eta) = |\eta|_g$, and let Φ_t be its Hamilton flow. Differentiating the eikonal equation shows that the stationary equations above describe exactly

$$(x,\eta) = \Phi_{\pm t}(y,\xi).$$

The plus branch has $\partial_t \phi_+ = -b$ and follows Φ_t; the minus branch follows Φ_{-t}. We obtain

$$\mathrm{WF}(A_\pm(t)f) \subset \Phi_{\pm t}\big(\mathrm{WF}(f)\big).$$

The parametrix and the exact solution differ by a smooth term, as we justify in the next section, so this comparison also applies to the exact solution on each branch.

For the half-wave propagators $U_\pm(t)$ constructed in Section 10, the comparison is an equality. Indeed, apply the same inclusion to the reverse propagator and use $U_\pm(-t)U_\pm(t) = \mathrm{Id}$ to get

$$\mathrm{WF}\big(U_\pm(t)h\big) = \Phi_{\pm t}\big(\mathrm{WF}(h)\big).$$

For the full wave equation we must keep both Cauchy data and separate the two branches; their contributions can cancel in $u(t,\cdot)$ alone. For longer times we repeat the short-time comparison along the rays. A caustic can make the phase representation fail, but the ray flow in the cotangent bundle remains smooth. Chapter III gives the detailed wave front argument and explains how Fourier integral operators continue the parametrix through such points.

6. Justification of the parametrix

Let us see what we have done so far. For every $(x_0, \xi^0) \in T^*\mathbf{R}^n \setminus 0$, we showed that one can solve the eikonal and the transport equations near x_0 and for $\xi = \xi^0$. By continuity, one can do that for ξ in some conic neighborhood of ξ^0.

Assume $(f_1, f_2) \in \mathcal{E}'(\mathbf{R}^n)$. Over the compact support of the data, a finite number of these patches cover the unit covectors. Use a pseudodifferential partition of unity there and cut off near $\xi = 0$. The omitted low frequencies have a smooth kernel. Summing the local constructions gives u_0 for $|t| \ll 1$ with

$$Pu_0 \in C^\infty, \quad u_0|_{t=0} - f_1 \in C^\infty, \quad \partial_t u_0|_{t=0} - f_2 \in C^\infty. \tag{6.1}$$

Why does this give a parametrix in the sense of Chapter I? Let u be the exact solution and set $w = u - u_0$. Then

$$Pw = -Pu_0, \quad w|_{t=0} = f_1 - u_0|_{t=0}, \quad \partial_t w|_{t=0} = f_2 - \partial_t u_0|_{t=0}. \tag{6.2}$$

The forcing and both initial data are smooth. Local inhomogeneous energy estimates, applied at every Sobolev order, give $w \in C^\infty$ on the same time interval. The residual and the errors in the two traces have smooth kernels. We may differentiate in the data variable and apply the same estimates. Thus the exact and approximate solution operators differ by smoothing operators.

7. High frequency waves

The parametrices constructed above aimed to construct solutions, up to smooth terms, with arbitrarily prescribed Cauchy data at $t = \text{const.}$. They can be used, as we will see later, to derive the propagation of singularities results for the wave equation. Sometimes we want to construct solutions which are smooth but highly oscillatory, depending on a large parameter, so that most of the energy propagates along the same geodesics as above.

We are looking for solutions of the type

$$u = e^{\mathrm{i}\lambda\phi(t,x,\omega)} a(t,x,\omega,\lambda),$$

where

$$a(t,x,\omega,\lambda) \sim \sum_{k=0}^{\infty} \lambda^{-k} a_k(t,x,\omega), \tag{7.1}$$

compare with (2.5), (2.6). Here, ω is a parameter determining the direction of the propagation (say, at $t = 0$) assumed to belong to a compact set.

We can think of λ as $|\xi|$ and ω as its direction. Here we keep the direction fixed and study an individual oscillatory wave as λ increases.

7.1. The Euclidean case. To have an explicit example, note that for the Euclidean wave equation, we can just take

$$u = e^{\mathrm{i}\lambda(t-x\cdot\omega)}, \quad |\omega| = 1.$$

We could have also replaced t by $-t$ there, corresponding to the two choices of phase functions in (2.5). This is an exact solution describing a "wave" propagating in the direction ω since we can write the phase function as $t - x\cdot\omega = -(x - t\omega)\cdot\omega$. It has uniform local energy and it is not concentrated near a certain line, as we may want. To achieve the latter, look at the Cauchy data at $t = 0$ first:

$$(u, u_t)|_{t=0} = (e^{-\mathrm{i}\lambda x\cdot\omega}, \mathrm{i}\lambda e^{-\mathrm{i}\lambda x\cdot\omega}).$$

They are not localized in space (not even asymptotically). We may want to do that. Let $\chi \in C_0^\infty(\mathbf{R}^n)$ be a cut-off function supported in a "small" neighborhood of some x_0. We would like to have a solution supported "near" the line $x = x_0 + t\omega$ with

$$u|_{t=0} = \chi(x)e^{-\mathrm{i}\lambda x\cdot\omega}, \tag{7.2}$$

and we will discuss $u_t|_{t=0}$ later. Now, we are looking for u of the form (7.1).

Lemma 2.1 here reduces to the following (with the exponents considered as multiplication factors), where $\Box = \partial_t^2 - \Delta$:

$$e^{-\mathrm{i}\lambda\phi}\Box e^{\mathrm{i}\lambda\phi} = \lambda^2(-\phi_t^2 + |\nabla_x\phi|^2) + 2\mathrm{i}\lambda(\phi_t\partial_t - \nabla_x\phi\cdot\nabla_x) + \mathrm{i}\lambda\Box\phi + \Box. \tag{7.3}$$

In our case, $\phi = t - x\cdot\omega$ solves the eikonal equation, so we get the transport equation

$$2\mathrm{i}\lambda(\partial_t + \omega\cdot\nabla_x)a + \Box a = 0, \quad a|_{t=0} = \chi(x).$$

Equate successive powers of λ in (7.1) to get

$$(\partial_t + \omega\cdot\nabla_x)a_0 = 0, \qquad a_0|_{t=0} = \chi(x), \tag{7.4}$$

$$(\partial_t + \omega\cdot\nabla_x)a_k = -\frac{1}{2\mathrm{i}}\Box a_{k-1}, \quad a_k|_{t=0} = 0, \quad k = 1, 2, \dots. \tag{7.5}$$

Those are just simplified versions of (4.2), (4.3). We have

$$a_0 = \chi(x - t\omega)$$

and a_k, $k \geq 1$ are non-trivial; for example. a_1 solves

$$(\partial_t + \omega\cdot\nabla_x)a_1 = -\frac{1}{2\mathrm{i}}\left[\left((\omega\cdot\nabla_x)^2 - \Delta\right)\chi\right](x - t\omega), \quad a_1|_{t=0} = 0.$$

The solution is just an integral of the r.h.s. along the characteristics of the fist order operator on the left. To simplify the formula, assume $\omega = (0, \dots, 0, 1)$, which we can only do by rotating the coordinates. Then we get

$$(\partial_t + \partial_{x^n})a_1 = \frac{1}{2\mathrm{i}}\Delta_{x'}\chi(x', x^n - t), \quad a_1|_{t=0} = 0,$$

where $x = (x', x^n)$ and $\Delta_{x'}$ is the Laplacian in the x' variables. Perform the change of variables $y^n = x^n - t$, keeping t and x' unchanged, to get

$$\partial_t a_1 = \frac{1}{2\mathrm{i}}\Delta_{x'}\chi(x', y^n), \quad a_1|_{t=0} = 0.$$

Therefore,

$$a_1 = \frac{t}{2\mathrm{i}}\Delta_{x'}\chi(x', x^n - t).$$

Then by (7.5), we need to solve

$$(\partial_t + \partial_{x^n})a_2 = -\frac{1}{2\mathrm{i}}\Box a_1 = -\frac{1}{4}t\Delta_{x'}^2\chi(x', x^n - t) - \frac{1}{2}\partial_{x^n}\Delta_{x'}\chi(x', x^n - t), \quad a_2|_{t=0} = 0.$$

Then

$$a_2 = -\frac{t^2}{8}\Delta_{x'}^2\chi(x', x^n - t) - \frac{t}{2}\partial_{x^n}\Delta_{x'}\chi(x', x^n - t).$$

We can then compute all the other a_k in this way.

We use (7.1) as an asymptotic expansion: after N terms the remainder in the amplitude is of order λ^{-N} on a fixed time interval. Let us see separately what error is left in the equation. First notice that if the cutoff depends only on $x \cdot \omega$, say $\chi(x) = F(x \cdot \omega)$, then $e^{\mathrm{i}\lambda(t-x\cdot\omega)}F(x \cdot \omega - t)$ is already an exact solution. For a cutoff in all the variables the transverse derivatives above give the corrections.

So far we have prescribed only the displacement. The initial velocity of this one-branch construction is $e^{-\mathrm{i}\lambda x\cdot\omega}(\mathrm{i}\lambda\chi(x) + O(1))$. The lower order terms in the velocity are determined by differentiating the amplitude at $t = 0$.

If we prescribe $u_t|_{t=0} = 0$ instead, we need both phases $t - x \cdot \omega$ and $-t - x \cdot \omega$ to keep the same initial oscillation. Their leading amplitudes are $\chi(x - t\omega)/2$ and $\chi(x + t\omega)/2$, respectively. Figure 6 shows a numerical example.



Figure 6. A localized high-frequency wave with zero initial velocity. The plots show $\Im u$ at $t = 0$, T and $2T$. The two packets have half the initial displacement amplitude and, to leading order, each carries half the initial energy.

On the first branch each $a_k(t, \cdot)$ is supported in $t\omega + \operatorname{supp}\chi$. Indeed, differentiation preserves this support and we integrate the next transport equation along the same lines. What does this tell us about an exact wave?

Fix $T > 0$ and $N \geq 1$. For $\lambda \geq 1$ put

$$v_N = e^{\mathrm{i}\lambda(t-x\cdot\omega)}\sum_{k=0}^{N-1}\lambda^{-k}a_k(t, x).$$

Let u solve $\Box u = 0$ with the two Cauchy data of v_N at $t = 0$. In particular its displacement is exactly (7.2). For this comparison the velocity is $\partial_t v_N(0,\cdot)$, including the lower order terms. By (7.5), all but the last term in the residual cancel. We get

$$\Box v_N = \lambda^{1-N} e^{\mathrm{i}\lambda(t-x\cdot\omega)} \Box a_{N-1}.$$

The function $\Box a_{N-1}$ is smooth and has uniformly bounded spatial support for $|t| \le T$. Its L^2 norm is therefore bounded by a constant $C_{N,T,\chi}$. Set $w = u - v_N$. Then $\Box w = -\Box v_N$ and both initial data vanish. Multiply the equation by $\overline{w_t}$, integrate in x and take real parts. For either sign of t the energy inequality gives

$$\Big(\|w_t(t)\|_{L^2}^2 + \|\nabla_x w(t)\|_{L^2}^2 \Big)^{1/2} \le \int_{\min(0,t)}^{\max(0,t)} \|\Box v_N(s)\|_{L^2}\, \mathrm{d}s \le C_{N,T,\chi} |t| \lambda^{1-N}.$$

Integrating w_t once also bounds $\|w(t)\|_{L^2}$. Thus, on any fixed time interval, keeping $N = M + 1$ terms gives an $O(\lambda^{-M})$ error in $H^1 \times L^2$ for these matched data. A discrepancy in the initial velocity contributes its L^2 norm to the energy bound. This is why specifying only its leading term does not give an estimate to arbitrary order.

7.2. The general case. For variable speed, take a real phase on a patch where the ray coordinates are regular. The same cancellation leaves $\lambda^{1-N} e^{\mathrm{i}\lambda\phi} P a_{N-1}$ after N terms. On a fixed time interval on which the transported support stays inside the patch, the inhomogeneous energy estimate gives the same order of error for matched data. Its constant now depends on the coefficients and the time interval. If the ray coordinates for this phase become singular, we must first change the representation of the wave.

8. Two parameter expansions; beams

In (7.2), χ controls the width of the wave packet at the initial moment. The high-frequency part of u, for a fixed finite time interval $|t| \le T$, propagates along the characteristics lines $x = x_0 + t\omega$ originating from points $x_0 \in \operatorname{supp} \chi$. We may want to introduce a second (small) parameter, $\varepsilon > 0$, controlling the width of that wave package or the width of the beam by replacing $\chi(x)$ by $\chi(x/\varepsilon)$. Then

$$u|_{t=0} = \chi(x/\varepsilon) e^{-\mathrm{i}\lambda x\cdot\omega}, \tag{8.1}$$

Take $0 < \varepsilon \le 1$. On a fixed time interval a_k involves at most $2k$ derivatives of the initial cutoff, by (7.5). Rescaling the cutoff therefore costs up to ε^{-2k} at order k. The successive orders in the amplitude are now bounded by constants times $(\lambda\varepsilon^2)^{-k}$. To get an asymptotic scale we need $\lambda\varepsilon^2 \to \infty$. For example, $\varepsilon = \lambda^{-\alpha}$ with $0 < \alpha < 1/2$ works. What happens at $\alpha = 1/2$? We take up this critical scale in Chapter IV.

9. The Schwartz kernel of the solution operator of the wave equation. A generalized Kirchhoff formula

In three dimensions the classical Kirchhoff formula expresses the solution of the Euclidean wave equation for $t > 0$ in terms of the data on the sphere $|x - y| = t$. In particular, if $f_1 = 0$ and $f_2 = \delta$, the forward solution is G_0, where

$$G_0(t,x) = \frac{1}{4\pi|x|} \delta(t - |x|).$$

If f_2 is a general function (or a distribution), then $u = G_0 * f_2$, where the convolution is in the x variable. The latter is the classical Kirchhoff formula when $f_1 = 0$, and the case $f_1 \neq 0$ easy follows from that one.

Let us derive the leading term for (1.1) in three dimensions. For y fixed, write δ_y for the point mass relative to $\mathrm{d}\operatorname{Vol}_{g_0}$. We want to solve

$$(\partial_t^2 - c^2\Delta_{g_0})G = 0, \quad G|_{t=0} = 0, \quad G_t|_{t=0} = \delta_y,$$

for $t \geq 0$ which is also equivalent to

$$(\partial_t^2 - c^2\Delta_{g_0})G = \delta(t)\delta_y, \quad G|_{t<0} = 0.$$

Let $\rho(x, y)$ be distance in the wave speed metric $g = c^{-2}g_0$ for the operator just displayed. Near the first arrival we look for the progressive wave expansion of Chapter I,

$$G(t, x, y) \sim \frac{1}{\rho(x, y)} \sum_{j=0}^{\infty} a_j(x, y)h_j(t - \rho(x, y)). \tag{9.1}$$

Here $h_0 = \delta$, h_1 is the Heaviside function and $h_{j+1}(s) = s_+^j/j!$ for $j \geq 1$. Fix y and set $r(x) = \rho(x, y)$. We work in a normal neighborhood of y, with $x \neq y$, so r is smooth there.

The distance solves $c^2|\nabla r|_{g_0}^2 = 1$, and $\mathrm{d}r$ is the unit covector dual in g to the final tangent of the minimizing geodesic. With the Fourier kernel $e^{\mathrm{i}\lambda t}$ from Chapter I, the retarded Euclidean kernel gives the outgoing Helmholtz kernel $e^{\mathrm{i}\lambda|x-y|}/(4\pi|x - y|)$.

Let $v = \exp_y^{-1} x$ for the metric g. In polar coordinates in T_yM write $r = |v|_{g(y)}$ and $\omega = v/r$. Then

$$g = \mathrm{d}r^2 + r^2h(r, \omega), \tag{9.2}$$

where $h(r, \omega)$ is a metric in the angular variables. For the rest of this calculation assume $c = 1$, so $g = g_0$.

In particular, if x' parameterizes S^2 locally, $x = (x', x^3 = r)$ are boundary normal coordinates at any geodesic sphere $r(x) = \text{const.}$ (where the coordinates are still valid) with x^3 shifted by that constant. In those coordinates,

$$\Delta_g = \frac{1}{r^2|h|^{1/2}}\partial_r\big(r^2|h|^{1/2}\partial_r\big) + \frac{1}{r^2}\Delta_h, \quad \Delta_h := \frac{1}{|h|^{1/2}}\partial_{x^\alpha}\big(|h|^{1/2}h^{\alpha\beta}\partial_{x^\beta}\big).$$

Apply $\partial_t^2 - \Delta_g$ to (9.1) and collect terms of the same order of singularity as in Chapter I. The coefficient of $\delta'(t - r)$ gives the first transport equation,

$$\left(\partial_r + \frac{1}{2}(\Delta_g r)\right)\frac{a_0}{r} = 0,$$

i.e.,

$$\left(\partial_r + \frac{1}{4}\partial_r \ln(r^4|h|)\right)\frac{a_0}{r} = 0.$$

Therefore,

$$a_0(r, \omega)|h(r, \omega)|^{1/4} = C(\omega).$$

How do we find $C(\omega)$? Let $g_{\mathbb{S}^2}$ be the round metric on the unit sphere in T_yM, expressed in the same angular coordinates. The Riemannian volume in these polar coordinates is $r^2\sqrt{|h(r, \omega)|}\,\mathrm{d}r\,\mathrm{d}\omega_{\mathrm{loc}}$,

where $\mathrm{d}\omega_{\mathrm{loc}}$ is coordinate measure. Dividing by Euclidean tangent-space volume in polar coordinates gives the volume Jacobian of $\exp_y$,

$$J(x,y) = \frac{\sqrt{|h(r,\omega)|}}{\sqrt{|g_{\mathbb{S}^2}(\omega)|}}, \qquad J(y,y) = 1.$$

Thus $a_0\sqrt{J}$ is constant along each ray. For $c = 1$ take the point source normalized w.r.t. Riemannian volume. In orthonormal normal coordinates at y the leading term must match the Euclidean point source, so $a_0 \to 1/(4\pi)$ in every direction. The 4π comes from integration over the unit sphere. We obtain

$$G(t,x,y) = \frac{\delta(t-\rho(x,y))}{4\pi\rho(x,y)\sqrt{J(x,y)}} + \text{smoother terms}.$$

Let us assume that we are solving the following inverse problem. We have an "earthquake" at $t = 0$ at position y and we are sitting at x and waiting to see how long it takes for the signal to arrive there. This happens at $t = \rho(x,y)$, we are measuring the geodesic distance between x and the epicenter y. This comes from the delta term above. Does the amplitude provide extra information? The latter gives us $J(x,y)$.

Suppose now that y and x vary on the boundary near the endpoints of a unique transverse minimizing geodesic without conjugate points. Tangential differentiation of the distance gives the tangential parts of the initial and final unit covectors, with opposite signs at the two ends. This suggests a question about the direct arrival for the ambient wave kernel. Under what assumptions does knowledge of the boundary distance near this pair determine the whole singular expansion of G near $t = \rho(x,y)$? The calculation above identifies the leading transport factor. Recovering it and the successive transport coefficients from boundary measurements alone requires a further argument.

10. "Half-wave equation". General hyperbolic ΨDO equations of real principal type

We have been using two phases to solve the wave equation. Can we separate the two branches at the operator level? Return to (1.1), with the global assumptions of Section 1, and write

$$L = -c^2\Delta_{g_0}, \qquad B = L^{1/2}.$$

Here L is the nonnegative self-adjoint operator in $L^2(\mathrm{d}\mu)$, $\mathrm{d}\mu = c^{-2}\mathrm{d}\,\mathrm{Vol}_{g_0}$, used in Section 1. The spectral theorem defines its nonnegative square root B. Since the coefficients are independent of t, we have the factorization

$$\partial_t^2 + L = (\partial_t + \mathrm{i}B)(\partial_t - \mathrm{i}B).$$

This suggests solving the two first order problems

$$(\partial_t \pm \mathrm{i}B)v_\pm = 0, \qquad v_\pm|_{t=0} = h_\pm.$$

Their solution operators are the two *half-wave propagators*,

$$v_\pm(t) = U_\pm(t)h_\pm, \qquad U_\pm(t) = e^{\mp\mathrm{i}tB}.$$

Indeed, differentiation gives $\partial_t U_\pm = \mp\mathrm{i}BU_\pm$, and $U_\pm(0) = \mathrm{Id}$. The spectral theorem also gives

$$U_\pm(t+s) = U_\pm(t)U_\pm(s), \qquad U_\pm(t)^{-1} = U_\pm(-t) = U_\mp(t).$$

Both groups are unitary in $L^2(\mathrm{d}\mu)$. Each $v_\pm$ solves the second order wave equation as well.

For $c = 1$ and g_0 Euclidean, $B = |D_x|$ is the Fourier multiplier $|\xi|$. Thus

$$U_\pm(t)h(x) = (2\pi)^{-n} \int e^{\mathrm{i}(x\cdot\xi \mp t|\xi|)} \hat{h}(\xi)\, \mathrm{d}\xi.$$

These are exactly the two phases (1.10). The plus branch has temporal covector $\tau = -|\xi|$, and its ray moves along $x = y + t\xi/|\xi|$.

Matching the Cauchy data. How do we choose h_+ and h_-? First suppose that the data are spectrally supported where $B \geq \delta > 0$, so B^{-1} is bounded on this subspace. For $u = U_+(t)h_+ + U_-(t)h_-$, the two initial conditions give

$$h_+ + h_- = f_1, \qquad -\mathrm{i}Bh_+ + \mathrm{i}Bh_- = f_2.$$

Solving this system, we get

$$h_\pm = \frac{1}{2}\big(f_1 \pm \mathrm{i}B^{-1}f_2\big), \qquad u(t) = \frac{1}{2}U_+(t)\big(f_1 + \mathrm{i}B^{-1}f_2\big) + \frac{1}{2}U_-(t)\big(f_1 - \mathrm{i}B^{-1}f_2\big).$$

For $B = |D_x|$ this is (1.8), including the signs of the two velocity amplitudes. A pure plus wave has initial velocity $f_2 = -\mathrm{i}Bf_1$, and a pure minus wave has $f_2 = \mathrm{i}Bf_1$.

What happens near spectral value zero? Combine the two terms before removing the spectral cutoff. We obtain

$$u(t) = \cos(tB)f_1 + s_t(B)f_2, \qquad s_t(\sigma) = \begin{cases} \sin(t\sigma)/\sigma, & \sigma > 0, \\ t, & \sigma = 0. \end{cases}$$

The function s_t is continuous and bounded on $[0, \infty)$ for fixed t. This gives the exact solution for $f_1 \in \mathcal{D}(B)$ and $f_2 \in L^2(\mathrm{d}\mu)$, including any zero spectral component. In the microlocal construction we work at large $|\xi|$ and use an elliptic parametrix for B^{-1}. The two-branch formula then holds modulo smoothing operators for compactly supported distribution data.

The square root and the rays. What kind of operator is B for variable c and g_0? Locally at nonzero covectors, elliptic functional calculus gives a classical ΨDO of order one with principal symbol

$$b(x,\xi) = c(x)|\xi|_{g_0} = |\xi|_g, \qquad g = c^{-2}g_0.$$

We can find its lower order symbol by requiring $B^2 = L$. Start with an operator having symbol b. The difference between its square and L has order one. Adding a symbol b_0 of order zero changes that leading error by $2b\,b_0$. Since $b > 0$ for $\xi \neq 0$, this determines b_0. Repeating the argument determines the terms of orders $-1, -2, \ldots$. Asymptotic summation gives the square root symbol modulo smoothing terms.

Now construct $U_\pm(t)$ for small $|t|$ on a conic patch. As in (5.1), we look for

$$U_\pm(t)h(x) = (2\pi)^{-n} \int e^{\mathrm{i}\phi_\pm(t,x,\xi)} A_\pm(t,x,\xi)\hat{h}(\xi)\, \mathrm{d}\xi \mod C^\infty.$$

Here $A_\pm$ is a classical symbol of order zero. At $t = 0$ take $\phi_\pm = x\cdot\xi$ and $A_\pm = 1$ at high frequencies. Substitution in the first order equation gives

$$\partial_t\phi_\pm \pm b(x, \partial_x\phi_\pm) = 0.$$

This is precisely (3.19). The lower order terms give transport equations along the same rays, with leading initial amplitude 1. Equivalently, the leading amplitude satisfies (4.2) for the wave operator

(1.1). Summing the transport expansion and correcting the smooth error gives the exact propagator microlocally.

Let Φ_t denote the Hamilton flow of b. The plus branch follows Φ_t, and the minus branch follows Φ_{-t}. In the notation of Section 5, the stationary condition in the kernel is

$$y = \partial_\xi \phi_\pm(t, x, \xi), \qquad \eta = \partial_x \phi_\pm(t, x, \xi).$$

Thus $(x, \eta) = \Phi_{\pm t}(y, \xi)$. This is the correspondence between initial and final covectors found in Section 5. Chapter III will interpret it as the canonical relation of a Fourier integral operator, with the details of the oscillatory integral argument. Applying the wave front estimate sketched in Section 5, we get

$$\mathrm{WF}(U_\pm(t)h) \subset \Phi_{\pm t}\big(\mathrm{WF}(h)\big)$$

for compactly supported distribution data. The reverse propagator has the same property and satisfies $U_\pm(-t)U_\pm(t) = \mathrm{Id}$. Apply it to the output, with local cutoffs as needed, to obtain the reverse inclusion. Hence

$$\mathrm{WF}(U_\pm(t)h) = \Phi_{\pm t}\big(\mathrm{WF}(h)\big).$$

This says that each singularity follows its ray, carrying its covector with it. For longer times we compose the short-time propagators using the group law. A caustic can force us to change the phase representation, while the correspondence between initial and final covectors continues smoothly.

A general first order equation. The same construction works locally for

$$(\partial_t + \mathrm{i}Q(t, x, D_x))v = 0, \qquad v|_{t=s} = h,$$

where Q is a scalar classical ΨDO of order one with real principal symbol $q(t, x, \xi)$. With $D_t = \frac{1}{\mathrm{i}}\partial_t$, the principal symbol of $D_t + Q$ is $\tau + q$. Its τ-derivative equals one, so each time slice is noncharacteristic for this equation. The eikonal equation and the ray equations are

$$\partial_t \phi + q(t, x, \partial_x \phi) = 0, \qquad \dot{x} = \partial_\xi q, \qquad \dot{\xi} = -\partial_x q.$$

The lower order symbol enters the transport equations. The same oscillatory integral construction gives the propagator from time s to time t. It pairs each initial covector at time s with the covector reached by these ray equations at time t. Taking $q = \pm|\xi|_g$ recovers our two half-waves.

More generally, near a characteristic point where a real scalar principal symbol $p(t, x, \tau, \xi)$ satisfies $\partial_\tau p \neq 0$, we can solve $p = 0$ for $\tau = -q(t, x, \xi)$. Microlocal factorization removes an elliptic factor and reduces the equation to this first order form, modulo smoothing terms. This is how the half-wave construction extends to real-principal-type equations at a simple root transverse to the time slices.

11. Helmholtz equations

Return to the wave equation (1.1) and write $L = -c^2\Delta_{g_0}$, as in Section 10. We first work locally, without a boundary. For a time-harmonic wave with $\lambda > 0$, we have

$$u(t, x) = e^{-\mathrm{i}\lambda t}v(x), \qquad (\partial_t^2 + L)u = e^{-\mathrm{i}\lambda t}(L - \lambda^2)v.$$

Thus the Helmholtz equation is

$$(L - \lambda^2)v = f, \qquad \text{or} \qquad (-c^2\Delta_{g_0} - \lambda^2)v = f,$$

if the wave source is $e^{-\mathrm{i}\lambda t}f(x)$. This agrees with the time Fourier transform used in Chapter I,

$$\widehat{u}(\lambda, x) = \int e^{\mathrm{i}\lambda t}u(t, x)\,\mathrm{d}t,$$

whose inverse uses $e^{-\mathrm{i}\lambda t}$. In the spacetime convention $e^{\mathrm{i}(t\tau+x\cdot\xi)}$, the temporal covector is therefore $\tau = -\lambda$. For the more general equation (2.2), with coefficients independent of t, the same substitution gives $(-c^2\Delta_{g_0,A} + q - \mathrm{i}\lambda a - \lambda^2)v = f$.

The outgoing sign. In the Euclidean case with constant c, the plane wave

$$e^{-\mathrm{i}\lambda t}e^{\mathrm{i}\lambda x\cdot\omega/c} = e^{\mathrm{i}\lambda(x\cdot\omega/c-t)}, \qquad |\omega| = 1,$$

moves in the direction $+\omega$. This is the plus half-wave branch: its temporal covector is negative and its spatial covector points along the ray.

The one-dimensional outgoing wave in Chapter V is $F(t - x/c)$ on $x > 0$. Taking $F(s) = e^{-\mathrm{i}\lambda s}$ gives the outgoing spatial factor $e^{\mathrm{i}\lambda x/c}$; $F(t + x/c)$ gives the incoming factor $e^{-\mathrm{i}\lambda x/c}$. Since the outward normal derivative at $x = 0$ is $-\partial_x$, their Dirichlet-to-Neumann multipliers are

$$\Lambda_{\mathrm{out}}(\lambda) = -\frac{\mathrm{i}\lambda}{c}, \qquad \Lambda_{\mathrm{in}}(\lambda) = +\frac{\mathrm{i}\lambda}{c}.$$

This is exactly (V.1.6) after replacing ∂_t by $-\mathrm{i}\lambda$. Here outgoing means moving from the boundary into the domain. It does not mean moving in the direction of the outward normal. If we used the time factor $e^{+\mathrm{i}\lambda t}$ instead, the spatial signs would reverse.

From the retarded wave to the resolvent. For the global statements, assume g_0 is Euclidean and $c = 1$ outside a compact set, as in Section 1. Use the same nonnegative self-adjoint operator L in $L^2(\mathrm{d}\mu)$ and put $B = L^{1/2}$. Let $S(t) = s_t(B)$ be the sine propagator of Section 10, with $s_t(\sigma) = \sin(t\sigma)/\sigma$ for $\sigma > 0$ and $s_t(0) = t$. Thus $S(t)f$ has initial displacement zero and initial velocity f. Extend it by zero to $t < 0$. The resulting retarded solution $G(t)f$ satisfies

$$(\partial_t^2 + L)G(t)f = \delta(t)f.$$

For $\Im z > 0$, exponential damping makes its time transform converge. The spectral theorem gives

$$R(z) := (L - z^2)^{-1} = \int_0^\infty e^{\mathrm{i}zt}S(t)\,\mathrm{d}t.$$

Indeed, on a spectral value σ of B, the integral is $(\sigma^2 - z^2)^{-1}$, including $\sigma = 0$ by continuity. This identity fixes the outgoing choice without a separate sign convention.

For compactly supported $f \in L^2(\mathrm{d}\mu)$ and $\lambda > 0$, the limiting absorption construction gives

$$R_{\mathrm{out}}(\lambda)f = \lim_{\varepsilon\downarrow 0}(L - (\lambda + \mathrm{i}\varepsilon)^2)^{-1}f, \qquad R_{\mathrm{in}}(\lambda)f = \lim_{\varepsilon\downarrow 0}(L - (\lambda - \mathrm{i}\varepsilon)^2)^{-1}f,$$

with convergence locally in H^2. These are not bounded inverses on unweighted L^2 at the continuous spectrum. The first is the transform of the retarded wave, interpreted by the damping just used; the second corresponds to the advanced wave. For real coefficients the two are related by complex conjugation.

In $\mathbf{R}^3$, with $c = 1$ and g_0 Euclidean, Section 9 gives

$$G(t, x, y) = \frac{\delta(t - |x - y|)}{4\pi|x - y|}.$$

Its time transform is therefore

$$R_{\mathrm{out}}(\lambda;x,y)=\frac{e^{\mathrm{i}\lambda|x-y|}}{4\pi|x-y|},\qquad R_{\mathrm{in}}(\lambda;x,y)=\frac{e^{-\mathrm{i}\lambda|x-y|}}{4\pi|x-y|}.$$

Both solve $(-\Delta_x-\lambda^2)R(\lambda;x,y)=\delta_y$. They differ in the direction in which the time-dependent spherical wave travels. This also explains the outgoing kernel used in (IV.1.3).

The radiation condition. In dimension $n\ge 2$, let $R=|x|$ and $\omega=x/|x|$ outside the compact perturbation. For a compactly supported source, the outgoing solution has the form

$$v(R\omega)=R^{-(n-1)/2}e^{\mathrm{i}\lambda R}\big(v_\infty(\omega)+O(R^{-1})\big).$$

The corresponding Sommerfeld condition is

$$v=O(R^{-(n-1)/2}),\qquad (\partial_R-\mathrm{i}\lambda)v=o(R^{-(n-1)/2}),\qquad R\to\infty,$$

uniformly in ω. The incoming condition has $\partial_R+\mathrm{i}\lambda$ instead. If the constant speed at infinity is c_∞, replace λ in these spatial formulas by λ/c_∞.

Under the stated assumptions, the outgoing condition selects a unique solution of $(L-\lambda^2)v=f$ for each $\lambda>0$. For uniqueness, apply Green's identity to a homogeneous outgoing solution on a large ball. Its imaginary part gives zero net flux through the sphere. The outgoing expansion then forces $v_\infty=0$. Rellich's uniqueness theorem for the Euclidean exterior equation gives $v=0$ there, and unique continuation gives $v=0$ everywhere. Trapping does not destroy this fixed-frequency uniqueness.

An incident plane wave is not itself subject to the outgoing radiation condition. For example, with $v_{\mathrm{inc}}=e^{\mathrm{i}\lambda x\cdot\omega}$, write

$$v=v_{\mathrm{inc}}+v_{\mathrm{sc}},\qquad (L-\lambda^2)v_{\mathrm{sc}}=-(L-\lambda^2)v_{\mathrm{inc}}.$$

The right-hand side is compactly supported. We require the scattered field v_{sc} to be outgoing and obtain it by applying $R_{\mathrm{out}}(\lambda)$ to that right-hand side.

Geometric optics at high frequency. Away from the source, seek a local solution in the form

$$v_\lambda=e^{\mathrm{i}\lambda\psi(x)}a(x,\lambda),\qquad a\sim\sum_{j=0}^{\infty}\lambda^{-j}a_j(x).$$

A direct calculation gives

$$\begin{aligned}e^{-\mathrm{i}\lambda\psi}(L-\lambda^2)e^{\mathrm{i}\lambda\psi}a=\lambda^2\big(c^2|\mathrm{d}\psi|_{g_0}^2-1\big)a\\ -\mathrm{i}\lambda c^2\big(2\langle\mathrm{d}\psi,\mathrm{d}a\rangle_{g_0}+(\Delta_{g_0}\psi)a\big)-c^2\Delta_{g_0}a.\end{aligned}$$

The leading term gives the spatial eikonal equation

$$c^2|\mathrm{d}\psi|_{g_0}^2=|\mathrm{d}\psi|_g^2=1,\qquad g=c^{-2}g_0.$$

This is (3.11), and $\phi(t,x)=\psi(x)-t$ solves the wave eikonal equation (2.11). Along its rays,

$$\dot x=c^2g_0^{-1}\mathrm{d}\psi,\qquad \frac{\mathrm{d}}{\mathrm{d}t}\psi(x(t))=1.$$

Thus ψ measures travel time, and $\psi(x)-t$ stays constant along the wave. The scaled spatial covector is $\mathrm{d}\psi$; the full covector of v_λ is $\lambda\,\mathrm{d}\psi$. At a fixed λ the Helmholtz operator is elliptic as a spatial differential operator. The ray description concerns the family as $\lambda\to+\infty$.

Put

$$T_\psi a = 2\langle \mathrm{d}\psi, \mathrm{d}a\rangle_{g_0} + (\Delta_{g_0}\psi)a.$$

Equating the successive powers of λ gives

$$T_\psi a_0 = 0, \qquad T_\psi a_{j+1} = \mathrm{i}\Delta_{g_0} a_j, \qquad j = 0, 1, \dots.$$

These are ODEs along the rays, since $c^2 T_\psi a = 2\,\mathrm{d}a/\mathrm{d}t + c^2(\Delta_{g_0}\psi)a$. Prescribe the amplitudes on a transverse hypersurface and integrate. Before a caustic, asymptotic summation gives a local solution with residual $O(\lambda^{-\infty})$ in every local C^k norm. For a real phase the leading transport equation also gives

$$\mathrm{div}_{g_0}\left(|a_0|^2 \,\mathrm{grad}_{g_0}\,\psi\right) = 0,$$

the conservation of leading flux through a ray tube. Lower order terms in the wave operator change the transport equations, as before.

There is still a distinction between a local construction and a global outgoing solution. The resolvent formula integrates the wave over all positive times. A short-time parametrix alone does not control all the returning rays. For a global outgoing approximation v_{app} with compactly supported residual, set

$$r_\lambda = (L - \lambda^2)v_{\mathrm{app}} - f, \qquad v = v_{\mathrm{app}} - R_{\mathrm{out}}(\lambda)r_\lambda.$$

Then v solves the equation exactly. To conclude that the correction is small as $\lambda \to \infty$, we need a resolvent estimate. For a compactly supported cutoff χ, a nontrapping metric gives

$$\|\chi R_{\mathrm{out}}(\lambda)\chi\|_{L^2(\mathrm{d}\mu)\to L^2(\mathrm{d}\mu)} \le C\lambda^{-1}, \qquad \lambda \ge \lambda_0 > 0.$$

Here nontrapping means that every geodesic of the travel-time metric escapes each compact set in both time directions. With trapping, the resolvent can grow much faster, so a small formal residual alone is not enough. We will use the same distinction when we turn to boundary value problems.

A related boundary application is given by Cekić, Krupchyk, Sahoo, and Uhlmann [1, Proposition 2.1 and Sections 3–4]. On a compact nontrapping Riemannian manifold with strictly convex boundary, they correct Gaussian beam quasimodes for $(-\Delta_g - \lambda^2)u = 0$ using a right inverse with an $O(\lambda^{-1})$ bound in L^2; this solvability estimate imposes no boundary condition. Near the boundary, the Gaussian quasimode splits into entering and exiting pieces. The center of the exiting piece locates the exit point, the tangential differential of its real phase gives the tangential exit covector, and its phase value, normalized to zero at entry, gives the geodesic travel time. A boundary integral identity and stationary phase then allow them to compare these lens data from high-frequency Dirichlet-to-Neumann maps.

Chapter III

Introduction to Fourier Integral Operators through geometric optics

Most of this chapter is AI generated after a series of prompts. The material is classical but I still need to review it. Sections 6 and 7 are new to me. Read at your own peril!

In Chapter II, Section 5 sketched how the parametrix compares wave front sets at two times, Section 6 justified the parametrix modulo a smooth remainder, and Section 10 introduced the two half-wave groups. We now return to these operators. We prove the local wave front estimate, derive the canonical relation, and introduce Fourier integral operators. Other initial phases and changes of phase will let us describe the same wave propagator through caustics.

1. The wave front set of the local solution

The parametrix (II.2.5) consists of terms of the type

$$
\begin{aligned}
u_0(t,x) &= (2\pi)^{-n} \int e^{\mathrm{i}\phi(t,x,\xi)} a(t,x,\xi) \hat{f}(\xi)\, \mathrm{d}\xi \\
&= (2\pi)^{-n} \iint e^{\mathrm{i}(\phi(t,x,\xi)-y\cdot\xi)} a(t,x,\xi) f(y)\, \mathrm{d}\xi\, \mathrm{d}y.
\end{aligned}
\tag{1.1}
$$

Here ϕ is real and homogeneous of degree one in ξ, and a is a symbol. We work on a small conic patch, cut off near $\xi = 0$, and take $f \in \mathcal{E}'$. For small $|t|$, our construction gives $\det \phi_{x\xi} \neq 0$ and $\partial_x \phi \neq 0$. We insert spatial cutoffs when needed. For fixed t, an operator of this form is our first local example of a *Fourier integral operator*, or FIO.

The wave front set of the local solution. Recall what it means for f to be smooth near a point and in a direction. We say that $(y_0, \xi_0) \notin \mathrm{WF}(f)$, where $\xi_0 \neq 0$, if there is $\rho \in C_0^\infty$, equal to

one near y_0, and an open cone Γ about ξ_0 such that

$$|\widehat{\rho f}(\xi)| \le C_M(1+|\xi|)^{-M}, \qquad \xi \in \Gamma, \quad M = 0, 1, \dots.$$

The complement of these points is the *wave front set*. Thus it records both the position and the covector direction of a singularity.

Let us test $u_0(t,\cdot)$ near x_0 in a direction η_0, with t fixed. Write $A_t f = u_0(t,\cdot)$ and multiply the output by a cutoff $\chi(x)$ near x_0. Localize the input by $\rho(y)$, and choose $\widetilde{\rho} \in C_0^\infty$ equal to one near $\operatorname{supp}\rho$. Since $\rho f = \widetilde{\rho}\,\rho f$, Fourier inversion of the localized datum gives

$$\widehat{\chi A_t(\rho f)}(\eta) = (2\pi)^{-2n} \iiiint e^{\mathrm{i}\Theta} b(x,y,\xi)\widehat{\rho f}(\zeta)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}\xi\,\mathrm{d}\zeta,$$

where

$$\Theta = \phi(t,x,\xi) - y\cdot\xi + y\cdot\zeta - x\cdot\eta, \qquad b = \chi(x)\widetilde{\rho}(y)a(t,x,\xi).$$

If a has order m, then b has compact support in (x,y) and satisfies

$$|\partial_x^\alpha \partial_y^\beta \partial_\xi^\gamma b| \le C_{\alpha,\beta,\gamma}(1+|\xi|)^{m-|\gamma|}.$$

Also, since ρf is a compactly supported distribution, there is some $q \ge 0$ such that

$$|\widehat{\rho f}(\zeta)| \le C(1+|\zeta|)^q.$$

We insert compact frequency cutoffs first, so all the integrations by parts below are justified. Their estimates will be uniform as the cutoffs are removed.

The derivatives of the phase in the variables x, y, ξ are

$$\Theta_x = \partial_x\phi - \eta, \qquad \Theta_y = \zeta - \xi, \qquad \Theta_\xi = \partial_\xi\phi - y.$$

Thus the stationary conditions are precisely

$$\eta = \partial_x\phi(t,x,\xi), \qquad \zeta = \xi, \qquad y = \partial_\xi\phi(t,x,\xi).$$

Where one of these derivatives is nonzero, we can integrate by parts using the corresponding operator

$$\mathcal{L}_x = \frac{\partial_x\phi - \eta}{\mathrm{i}|\partial_x\phi - \eta|^2}\cdot\partial_x, \qquad \mathcal{L}_y = \frac{\zeta - \xi}{\mathrm{i}|\zeta-\xi|^2}\cdot\partial_y,$$
$$\mathcal{L}_\xi = \frac{\partial_\xi\phi - y}{\mathrm{i}|\partial_\xi\phi - y|^2}\cdot\partial_\xi.$$

Each satisfies $\mathcal{L}e^{\mathrm{i}\Theta} = e^{\mathrm{i}\Theta}$. Moving $\mathcal{L}$ onto the amplitude means applying its formal transpose, including derivatives of its coefficients. None of these operators differentiates $\widehat{\rho f}(\zeta)$.

We first reduce to comparable frequencies. By homogeneity and $\partial_x\phi \ne 0$ on our compact conic patch,

$$c_0|\xi| \le |\partial_x\phi(t,x,\xi)| \le C_0|\xi|.$$

Consequently, outside a region

$$|\xi| \asymp |\eta|, \qquad |\zeta| \asymp |\xi|,$$

at least one of $|\partial_x\phi - \eta|$ and $|\zeta - \xi|$ is bounded below by a positive constant times $|\eta| + |\xi| + |\zeta|$. Here $\asymp$ means bounds in both directions by fixed positive constants, chosen sufficiently wide for this assertion. A smooth partition separates the two cases. Use $\mathcal{L}_x$ in the first case and $\mathcal{L}_y$ in

the second. Each integration by parts gains a factor of $(1+|\eta|+|\xi|+|\zeta|)^{-1}$. Derivatives of the coefficients have the same bounds. After N integrations, the absolute integrand is bounded by

$$C_N(1+|\eta|+|\xi|+|\zeta|)^{-N+\max(m,0)+q}.$$

Integrating in ξ,ζ therefore gives rapid decay in η if N is chosen large enough. Bounded frequencies cause no problem. We can henceforth restrict to $|\xi| \asymp |\eta| \asymp |\zeta|$ with $|\eta| \geq 1$.

Now stay away from the simultaneous stationary conditions in this comparable-frequency region. On each of finitely many smaller patches, one of the following holds for some fixed $\delta > 0$:

$$|\partial_x\phi - \eta| \geq \delta|\xi|, \qquad |\zeta - \xi| \geq \delta|\xi|, \qquad |\partial_\xi\phi - y| \geq \delta.$$

The cutoffs can be chosen homogeneous of degree zero in the frequency variables at large frequencies. Their ξ-derivatives therefore satisfy the usual symbol estimates. For the first two inequalities, $\mathcal{L}_x$ or $\mathcal{L}_y$ gains $|\xi|^{-1}$ at each step.

The third inequality needs a little care: $\partial_\xi\phi - y$ has degree zero, so the coefficients of $\mathcal{L}_\xi$ are only of order zero. But the derivative now falls on a symbol in ξ. In fact,

$${}^t\mathcal{L}_\xi b = -\frac{1}{\mathrm{i}}\partial_\xi \cdot \left(\frac{\partial_\xi\phi - y}{|\partial_\xi\phi - y|^2}\, b\right).$$

A ξ-derivative of the coefficient has order -1, and a ξ-derivative of b lowers its order by one. Thus ${}^t\mathcal{L}_\xi b$ has order $m-1$. Repeating the argument gives order $m-N$ after N integrations. The same statement holds for the first two operators on their respective patches. Derivatives falling on the auxiliary frequency cutoffs satisfy these bounds as well.

Let $I_N(\eta)$ be the contribution of one such nonstationary patch after these integrations. The (x,y) integration is over a fixed compact set, and the comparable ξ and ζ regions have volumes $O(|\eta|^n)$ each. We obtain the explicit bound

$$|I_N(\eta)| \leq C_N|\eta|^{m+q+2n-N}.$$

Given any M, taking $N > m+q+2n+M$ gives $I_N(\eta) = O(|\eta|^{-M})$. This proves rapid decay for every contribution away from the stationary set, without any smoothness assumption on f.

It remains to consider a neighborhood of a stationary point. If its input covector (y_0,ξ_0) is outside $\mathrm{WF}(f)$, choose ρ equal to one near y_0, with sufficiently small support, and an input cone Γ so that

$$|\widehat{\rho f}(\zeta)| \leq C_N(1+|\zeta|)^{-N}, \qquad \zeta \in \Gamma,$$

for every N. At the stationary point $\zeta = \xi$, so we may restrict to a smaller neighborhood with $\zeta \in \Gamma$. Here no integration by parts is needed: absolute values give

$$|I(\eta)| \leq C_N|\eta|^{m+2n-N}, \qquad |\xi| \asymp |\eta| \asymp |\zeta|.$$

This is again rapidly decreasing. The complementary patches are covered by the nonstationary argument.

Suppose now that no input covector in $\mathrm{WF}(f)$ satisfies the stationary conditions for (x_0,η_0). After shrinking χ and the output cone about η_0, a finite partition of the compact normalized stationary set reduces us to the regular-input patches just considered. All remaining pieces are nonstationary. Summing their estimates gives

$$|\widehat{\chi u_0(t,\cdot)}(\eta)| \leq C_M(1+|\eta|)^{-M}$$

for every M in that output cone. This is the definition of $(x_0, \eta_0) \notin \mathrm{WF}(u_0(t,\cdot))$. Let us state the result, including its spacetime version.

Theorem 1.1 (Wave front set of the local oscillatory integral)**.** *Let $f \in \mathcal{E}'$ and let u_0 be given by (1.1), with a real phase ϕ homogeneous of degree one in ξ and a symbol a, both smooth in t. Work on a conic patch where $\det \phi_{x\xi} \neq 0$ and $\partial_x \phi \neq 0$, with the spatial and low-frequency cutoffs described above. For each fixed t, we have, microlocally on this patch,*

$$\mathrm{WF}\big(u_0(t,\cdot)\big) \subset \big\{(x, \partial_x\phi(t,x,\xi))\,;\, (\partial_\xi\phi(t,x,\xi),\xi) \in \mathrm{WF}(f)\big\}.$$

For the wave front set in spacetime, we have

$$\mathrm{WF}(u_0) \subset \big\{(t,x,\partial_t\phi(t,x,\xi),\partial_x\phi(t,x,\xi))\,;\, (\partial_\xi\phi(t,x,\xi),\xi) \in \mathrm{WF}(f)\big\}.$$

Proof. The fixed-time statement was proved above. For the spacetime statement, include a cutoff in t and the factor $e^{-\mathrm{i}t\tau}$ in the Fourier transform. The additional stationary condition is $\tau = \partial_t\phi$. For integration by parts in (t,x), use the phase gradient $(\partial_t\phi - \tau, \partial_x\phi - \eta)$ and compare $|\xi|$ with $|(\tau,\eta)|$. The remaining steps are the same. □

These are inclusions: an amplitude can vanish, and different contributions can cancel. For a half-wave propagator the inverse gives equality; see Theorem 2.1 below. For either wave phase, these covectors are characteristic by the eikonal equation.

2. The canonical relation of the wave propagator

Keep t fixed. The kernel of (1.1) has phase

$$\Psi(t,x,y,\xi) = \phi(t,x,\xi) - y\cdot\xi.$$

The integration-by-parts argument gives the stationary set and the possible kernel covectors:

$$\Sigma_t = \{(x,y,\xi)\,;\; y = \partial_\xi\phi(t,x,\xi),\ \xi \neq 0\}, \qquad (\eta,\nu) = (\partial_x\phi, -\xi).$$

Here η is the kernel's x-covector and ν its y-covector. To read this as an operator acting on f, we must reverse the sign of the latter. Indeed, an input oscillation $e^{\mathrm{i}y\cdot\zeta}$ adds $y\cdot\zeta$ to the phase, and stationarity in y gives $\nu + \zeta = 0$. Thus the input covector is $\zeta = \xi$. The resulting relation between output and input is

$$C_t = \big\{(x, \partial_x\phi(t,x,\xi); y, \xi)\,;\, y = \partial_\xi\phi(t,x,\xi),\ \xi \neq 0\big\}.$$

This is the *canonical relation* associated with the local operator. It depends on the phase, independently of which datum we apply the operator to. The amplitude may remove singularities from this set, so the actual kernel wave front set need not fill it. With $C_t \circ \mathrm{WF}(f)$ denoting the possible outputs for inputs in $\mathrm{WF}(f)$, our estimate reads

$$\mathrm{WF}\big(u_0(t,\cdot)\big) \subset C_t \circ \mathrm{WF}(f).$$

Let us first check that C_t is a smooth manifold. The stationary equation determines y explicitly. Hence (x,ξ) are parameters, with parametrization

$$\iota_t(x,\xi) = \big(x, \partial_x\phi(t,x,\xi); \partial_\xi\phi(t,x,\xi), \xi\big).$$

Its differential has rank $2n$: the first and last components already contain x and ξ. In fact, projecting back onto these components is an inverse on its image. Thus C_t has dimension $2n$. Homogeneity gives

$$\partial_x\phi(t,x,s\xi) = s\,\partial_x\phi(t,x,\xi), \qquad \partial_\xi\phi(t,x,s\xi) = \partial_\xi\phi(t,x,\xi), \qquad s>0.$$

Therefore C_t is conic: simultaneous scaling of the output and input covectors leaves the base points fixed and keeps us in C_t.

The Lagrangian property. On the output and input cotangent bundles use the standard symplectic forms

$$\omega_X = \sum_{j=1}^{n} \mathrm{d}\eta_j \wedge \mathrm{d}x^j, \qquad \omega_Y = \sum_{j=1}^{n} \mathrm{d}\xi_j \wedge \mathrm{d}y^j.$$

The relevant form on their product is the difference

$$\Omega = \omega_X - \omega_Y = \mathrm{d}\big(\eta\cdot\mathrm{d}x - \xi\cdot\mathrm{d}y\big).$$

A submanifold is called *Lagrangian* if this form vanishes on every pair of its tangent vectors and its dimension is half the ambient dimension. Here the ambient dimension is $4n$, and we have already checked that $\dim C_t = 2n$. We only need the vanishing.

All differentials in the next calculation are in (x,ξ), with t fixed. Substituting $y = \partial_\xi\phi$ and $\eta = \partial_x\phi$, we have

$$\mathrm{d}\phi = \eta\cdot\mathrm{d}x + y\cdot\mathrm{d}\xi, \qquad \mathrm{d}(y\cdot\xi) = \xi\cdot\mathrm{d}y + y\cdot\mathrm{d}\xi.$$

Subtracting gives

$$\iota_t^*\big(\eta\cdot\mathrm{d}x - \xi\cdot\mathrm{d}y\big) = \mathrm{d}_{x,\xi}\big(\phi - \xi\cdot\partial_\xi\phi\big) = 0.$$

The star denotes pullback: we express the form in the parameters (x,ξ). The last equality is Euler's identity for a function homogeneous of degree one, $\phi = \xi\cdot\partial_\xi\phi$. Exterior differentiation commutes with this substitution, so

$$\iota_t^*\Omega = \mathrm{d}_{x,\xi}\left(\iota_t^*(\eta\cdot\mathrm{d}x - \xi\cdot\mathrm{d}y)\right) = 0.$$

This proves the Lagrangian property directly from the phase.

There is an additional property in our short-time construction. Since $\det\phi_{x\xi} \neq 0$, the map $(x,\xi)\mapsto(y,\xi)$ has invertible Jacobian:

$$\det\frac{\partial(y,\xi)}{\partial(x,\xi)} = \det\phi_{\xi x} \neq 0.$$

Thus $y = \partial_\xi\phi$ determines x from (y,ξ) locally. Likewise $(x,\xi)\mapsto(x,\partial_x\phi)$ is locally invertible. We can therefore write C_t as the graph of a local diffeomorphism

$$F_t : (y,\xi) \longmapsto (x,\eta).$$

On this graph, the vanishing of $\omega_X - \omega_Y$ says precisely

$$F_t^*\omega_X = \omega_Y.$$

A diffeomorphism preserving the symplectic form is called a *canonical transformation*. This explains the name of the relation. The determinant condition gives the graph property; the Lagrangian calculation above used only the phase parametrization and homogeneity.

Identification with the ray flow. For the wave phases $\phi_\pm$, put $b(x,\eta) = |\eta|_g$. The eikonal equations are

$$\partial_t\phi_\pm \pm b(x,\partial_x\phi_\pm) = 0.$$

Here ξ is the fixed parameter labeling the initial covector, while $\eta = \partial_x \phi_\pm$ is the covector at time t. Differentiate in ξ to get

$$\big(\partial_t \pm b_\eta(x, \partial_x\phi_\pm) \cdot \partial_x\big)\partial_\xi \phi_\pm = 0.$$

Along a curve with $\dot{x} = \pm b_\eta(x, \eta)$, the value of $\partial_\xi \phi_\pm$ is constant. Since $\phi_\pm(0, x, \xi) = x \cdot \xi$, this constant is the starting point y.

To follow the other covector, differentiate the eikonal equation in x and apply the chain rule to $\eta(t) = \partial_x \phi_\pm(t, x(t), \xi)$. The terms containing $\phi_{\pm,xx} b_\eta$ cancel, leaving

$$\dot{x} = \pm b_\eta(x, \eta), \qquad \dot{\eta} = \mp b_x(x, \eta), \qquad (x(0), \eta(0)) = (y, \xi).$$

These are Hamilton's equations for $\pm b$, obtained also from (II.3.17) by using t as the parameter. Combining this identification with the wave front estimate gives the following version of Hörmander's propagation of singularities theorem.

Theorem 2.1 (Propagation of the wave front set)**.** *Let $U_\pm(t) = e^{\mp itB}$ be the half-wave propagators of Chapter II, Section 10, under the global assumptions made there. Let Φ_t be the Hamilton flow of the principal symbol $b(x, \xi) = |\xi|_g$ of B. The canonical relation of $U_\pm(t)$ is*

$$C_\pm(t) = \big\{(x, \eta; y, \xi)\,;\, (x, \eta) = \Phi_{\pm t}(y, \xi),\ \xi \neq 0\big\}.$$

For every $f \in \mathcal{E}'$ and every fixed t, we have the equality

$$\mathrm{WF}\big(U_\pm(t) f\big) = \Phi_{\pm t}\big(\mathrm{WF}(f)\big).$$

Proof. For small $|t|$, the calculation above identifies the canonical relation with the graph of $\Phi_{\pm t}$. The exact propagator and its parametrix differ by a smoothing operator, so Theorem 1.1 gives

$$\mathrm{WF}\big(U_\pm(t) f\big) \subset \Phi_{\pm t}\big(\mathrm{WF}(f)\big).$$

Now use time reversibility. The same local estimate applies to $U_\pm(-t)$, with spatial cutoffs near the covectors under consideration. Since $U_\pm(-t)U_\pm(t) = \mathrm{Id}$, we get

$$\mathrm{WF}(f) = \mathrm{WF}\big(U_\pm(-t)U_\pm(t) f\big) \subset \Phi_{\mp t}\big(\mathrm{WF}(U_\pm(t) f)\big).$$

Applying $\Phi_{\pm t}$ gives the reverse inclusion, since $\Phi_{\pm t} \circ \Phi_{\mp t} = \mathrm{Id}$. This proves equality for small times. For any fixed finite time, subdivide the time interval and repeat the short-time argument, using the group law for $U_\pm$ and the flow law for Φ. The same composition identifies the canonical relation with the graph of $\Phi_{\pm t}$. A caustic does not affect this argument: the flow remains smooth in the cotangent bundle. We describe the changes of phase needed there in Section 5. □

The equality is for each half-wave branch separately. Singularities of the two branches can cancel in their sum at a fixed time.

For the Euclidean phases $\phi_\pm = x \cdot \xi \mp t|\xi|$, the ray correspondence is

$$(x, \eta) = \left(y \pm t\frac{\xi}{|\xi|}, \xi\right).$$

The plus branch has temporal covector $-|\xi|$ and moves in the direction $+\xi/|\xi|$. At $t = 0$ the relation is the diagonal $(x, \eta) = (y, \xi)$, as for a pseudodifferential operator.

We can include time in the output variables as well. Since $\tau = \partial_t \phi_\pm = \mp b(x, \eta)$, the spacetime relation is

$$C_\pm = \big\{(t, x, \mp b(x, \eta), \eta; y, \xi)\,;\, (x, \eta) = \Phi_{\pm t}(y, \xi),\ \xi \neq 0\big\}.$$

It has parameters (t, x, ξ) and dimension $2n + 1$, half the dimension of $T^*(\mathbf{R}_t \times X) \times T^*Y$. The same calculation, now including $\mathrm{d}t$, gives

$$\iota_\pm^*(\tau\,\mathrm{d}t + \eta \cdot \mathrm{d}x - \xi \cdot \mathrm{d}y) = \mathrm{d}_{t,x,\xi}\big(\phi_\pm - \xi \cdot \partial_\xi \phi_\pm\big) = 0,$$

where $\iota_\pm$ is this parametrization. Hence $C_\pm$ is Lagrangian for $\mathrm{d}\tau \wedge \mathrm{d}t + \omega_X - \omega_Y$. The temporal covector is needed for this spacetime statement.

The full wave equation. We now complete the argument sketched in Chapter II, Section 5, using the same parametrix with lower order terms. There we obtained $\mathrm{WF}(A_\pm(t)f) \subset \Phi_{\pm t}(\mathrm{WF}(f))$ for a single oscillatory term. For the full solution, first combine the contributions of f_1 and f_2 with the same phase into $v_\pm$; the data carried by these branches are their initial traces $h_\pm$, defined below. We will obtain equality for each branch. Their opposite temporal covectors prevent cancellation in spacetime; at a fixed time we keep both Cauchy data.

Theorem 2.2 (Propagation for the wave equation)**.** *Let u solve*

$$Pu = 0, \qquad P = \partial_t^2 + a\partial_t - c^2\Delta_{g_0,A} + q, \qquad u(0) = f_1, \quad u_t(0) = f_2,$$

with $f_1, f_2 \in \mathcal{E}'$. Retain the global assumptions on c, g_0 of Chapter II; the coefficients a, A, q may be smooth, complex, and time dependent, with all their derivatives bounded on each finite time strip. Write the parametrix (II.2.5) as $v_+ + v_-$, continued by short-time parametrices, and put $h_\pm = v_\pm(0)$. These are the full branch traces, including all lower order amplitudes. They have the form

$$h_\pm = H_{1,\pm} f_1 + H_{2,\pm} f_2,$$

where $H_{1,\pm}$ and $H_{2,\pm}$ are pseudodifferential operators of orders 0 and -1, with principal symbols $1/2$ and $\pm\mathrm{i}/(2|\xi|_g)$, respectively. As in Section II.5, let $b(x,\xi) = |\xi|_g$ and let Φ_t be its Hamilton flow. Then

$$\mathrm{WF}(u) = \bigcup_{\pm} \big\{(t, x, \mp|\eta|_{g(x)}, \eta)\,;\ t \in \mathbf{R},\ (x,\eta) = \Phi_{\pm t}(y,\xi), \quad (y,\xi) \in \mathrm{WF}(h_\pm)\big\}. \tag{2.1}$$

In particular, $\mathrm{WF}(u)$ is a union of maximally extended null bicharacteristics of $p = -\tau^2 + |\xi|_g^2$. For each fixed t,

$$\mathrm{WF}(u(t)) \cup \mathrm{WF}(u_t(t)) = \Phi_t(\mathrm{WF}(h_+)) \cup \Phi_{-t}(\mathrm{WF}(h_-)). \tag{2.2}$$

Locally, only smoothness of the coefficients is needed, and these conclusions hold along the portions of the rays covered by the Cauchy evolution.

Proof. The operators $H_{j,\pm}$ are obtained by setting $t = 0$ in the parametrix; their principal symbols are the initial amplitudes (II.4.4), including the factor b^{-1} for the velocity datum. The matching of both Cauchy data determines their lower order terms. The justification of the parametrix gives $u = v_+ + v_-$ modulo a smooth function of (t, x).

The operators $A_\pm(t)$ in Section II.5 need not be invertible for an arbitrary amplitude. To get the reverse inclusion, normalize the same transport construction on one branch. Work first on a short-time conic patch, with all cutoffs equal to one near the covectors under consideration. Write

$$E_\pm(t,s)h(x) = (2\pi)^{-n} \iint e^{\mathrm{i}(\phi_\pm(t,s,x,\xi) - y\cdot\xi)} e_\pm(t,s,x,\xi) h(y)\,\mathrm{d}y\,\mathrm{d}\xi,$$

where $\phi_\pm(s,s,x,\xi) = x \cdot \xi$ and $\partial_t\phi_\pm \pm b(x, \partial_x\phi_\pm) = 0$.

Choose the full initial amplitude $e_\pm(s,s,x,\xi) \sim 1$: its leading term is one and all its negative-order terms are zero. Solving the transport equations to all orders gives

$$PE_\pm(t,s) \in \Psi^{-\infty}, \qquad E_\pm(s,s) = \mathrm{Id} \mod \Psi^{-\infty},$$

with smooth dependence of the smoothing remainders on both times.

Why does the displacement trace determine this construction? Once the branch and phase are fixed, the transport equations (II.4.2)–(II.4.3) are first order along the rays, transverse to every time slice since $\partial_t \phi_\pm \neq 0$. If two amplitudes give solutions modulo smoothing and have the same full trace amplitude at a time slice, their leading difference solves the homogeneous transport equation with zero initial value, so it vanishes. The next difference then has zero source and zero initial value, and also vanishes. Induction gives equality at every symbol order, hence a smoothing difference. This is uniqueness within one branch. It does not assert uniqueness for the second order Cauchy problem from displacement alone: the velocity trace on this branch is already determined by the transport equations.

Now fix three nearby times t, s, r and compare

$$F_\pm(t) = E_\pm(t,s)E_\pm(s,r) \quad \text{with} \quad E_\pm(t,r).$$

We must first check that the product is on the same branch. Its kernel has phase

$$\phi_\pm(t,s,x,\eta) - z\cdot\eta + \phi_\pm(s,r,z,\xi) - y\cdot\xi.$$

Stationarity in the intermediate variables (z,η) gives

$$z = \partial_\eta \phi_\pm(t,s,x,\eta), \qquad \eta = \partial_z \phi_\pm(s,r,z,\xi).$$

For nearby times these equations have a unique nondegenerate solution on our patch. Indeed, after writing $\xi = \lambda\omega$, $\eta = \lambda\theta$, the Hessian of the phase divided by λ in (z,θ) at $t = s = r$ is

$$\begin{pmatrix} 0 & -\mathrm{Id} \\ -\mathrm{Id} & 0 \end{pmatrix}.$$

It remains invertible for nearby times. Stationary phase therefore reduces the product to one oscillatory integral with a full classical amplitude. The critical value of the phase is $\phi_\pm(t,r,x,\xi) - y\cdot\xi$: its t-derivative and x-gradient satisfy the same eikonal equation, and at $t = s$ it equals $\phi_\pm(s,r,x,\xi) - y\cdot\xi$. Uniqueness of the eikonal equation identifies the phases. Equivalently, the stationary equations concatenate the two pieces of the same ray.

Next, P acts on the first factor only, so

$$PF_\pm(t) = \big(PE_\pm(t,s)\big)E_\pm(s,r) \in \Psi^{-\infty}, \qquad F_\pm(s) = E_\pm(s,r) \mod \Psi^{-\infty}.$$

Here the localized compositions with smoothing kernels are smoothing. Thus $F_\pm(t)$ and $E_\pm(t,r)$ solve the same equation modulo smoothing, have the same phase, and have the same displacement trace at $t = s$. In this short-time representation $\det \phi_{x\xi} \neq 0$, so equality of the trace kernels modulo smoothing is equality of their full trace amplitudes modulo symbols of order $-\infty$. The transport uniqueness just proved gives

$$E_\pm(t,s)E_\pm(s,r) = E_\pm(t,r) \mod \Psi^{-\infty}. \tag{2.3}$$

This uses the full stationary phase and transport expansions; matching only the leading amplitudes would not give a smoothing remainder.

The same uniqueness, applied to each of the two data operators defining $v_\pm$, gives $v_\pm(t) = E_\pm(t,0)h_\pm$ modulo smooth terms. Taking the two endpoints equal in (2.3) gives both inverse identities,

$$E_\pm(s,t)E_\pm(t,s) = \mathrm{Id} \mod \Psi^{-\infty},$$
$$E_\pm(t,s)E_\pm(s,t) = \mathrm{Id} \mod \Psi^{-\infty}.$$

The reverse parametrix uses the same branch, with the two times interchanged; its ray flow is $\Phi_{\pm(s-t)}$. For longer times, subdivide the interval and compose the short-time parametrices. The inverse is the product of their inverses in reverse order. The local composition identity makes this continuation independent of the subdivision modulo smoothing, on the ray patches in question. It remains valid through caustics, using short-time phase charts.

The wave front estimate applied in both directions, as in the proof of Theorem 2.1, therefore gives

$$\mathrm{WF}(v_\pm(t)) = \Phi_{\pm t}(\mathrm{WF}(h_\pm)).$$

The lower order coefficients change the transport equations, but not the phases or the ray flow.

The spacetime estimate in Theorem 1.1 gives

$$\mathrm{WF}(v_\pm) \subset \big\{(t,x,\mp b(x,\eta),\eta)\,;\,(x,\eta) \in \Phi_{\pm t}(\mathrm{WF}(h_\pm))\big\}.$$

To prove the reverse inclusion, fix t_0. The conormal directions to this time slice are $(t_0,x,\tau,0)$, $\tau \neq 0$, and the preceding inclusion excludes them. Thus restriction is defined and satisfies

$$\mathrm{WF}(v_\pm(t_0)) \subset \big\{(x,\eta)\,;\ (t_0,x,\tau,\eta) \in \mathrm{WF}(v_\pm) \text{ for some } \tau\big\}.$$

Indeed, after localizing near (t_0,x_0) and calling the result w, we have

$$\widehat{w(t_0,\cdot)}(\eta) = \frac{1}{2\pi}\int e^{\mathrm{i}t_0\tau}\widehat{w}(\tau,\eta)\,\mathrm{d}\tau.$$

The conormal condition gives rapid decay for $|\tau| > C|\eta|$. Outside the set on the right of the restriction estimate, compactness of the remaining normalized directions gives rapid decay also for $|\tau| \leq C|\eta|$, after shrinking the cutoff and the η-cone. Integration proves the inclusion. Restriction alone need not give equality.

The fixed-time equality supplies a spacetime covector over every point of $\Phi_{\pm t_0}(\mathrm{WF}(h_\pm))$. The spacetime inclusion permits only $\tau = \mp b(x,\eta)$, proving equality for each branch. Since $b > 0$ at nonzero spatial covectors, the two spacetime wave front sets are disjoint and cannot cancel in their sum. This proves (2.1). On $\tau = \mp b$, Hamilton's equations for p, parametrized by t, are

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \pm b_\xi, \qquad \frac{\mathrm{d}\xi}{\mathrm{d}t} = \mp b_x, \qquad \frac{\mathrm{d}\tau}{\mathrm{d}t} = 0.$$

Thus these curves are the null bicharacteristics, followed in both time directions.

It remains to recover the two branches at a fixed time. Set

$$K_\pm(t) = \partial_r E_\pm(r,t)|_{r=t}\,.$$

At $r = t$ the phase is $(x-y)\cdot\xi$, so $K_\pm(t)$ are spatial pseudodifferential operators of order one with principal symbols $\mp\mathrm{i}b$. Differentiating the composition identity gives $\partial_t v_\pm = K_\pm(t)v_\pm$ modulo

smooth terms. The difference $K_+ - K_-$ is elliptic, with principal symbol $-2ib$. If $Q(t)$ is its parametrix, then

$$\begin{aligned} v_+(t) &= Q(t)\big(u_t(t) - K_-(t)u(t)\big), \\ v_-(t) &= u(t) - v_+(t), \end{aligned} \qquad \text{mod } C^\infty. \tag{2.4}$$

Conversely, $u = v_+ + v_-$ and $u_t = K_+v_+ + K_-v_-$ modulo smooth terms. Pseudolocality in both directions gives

$$\mathrm{WF}(u(t)) \cup \mathrm{WF}(u_t(t)) = \mathrm{WF}(v_+(t)) \cup \mathrm{WF}(v_-(t)),$$

which proves (2.2). □

For $\mathrm{WF}(u(t))$ alone there is generally only inclusion in (2.2): $u(0) = 0$ can be smooth while $u_t(0)$ is singular. When $a = A = q = 0$, the branch data reduce to $h_\pm = \frac{1}{2}(f_1 \pm \mathrm{i}B^{-1}f_2)$ modulo smooth terms, with $B = \sqrt{-c^2\Delta_{g_0}}$. With lower order terms present, the full trace operators $H_{j,\pm}$ must be retained. Their principal symbols alone do not determine $\mathrm{WF}(h_\pm)$, since a correction of negative order need not be smoothing.

3. Other phase functions

We now replace $x \cdot \xi$ in the initial condition for the eikonal equation by a real function $\psi(x,\theta)$, homogeneous of degree one in $\theta \in \mathbf{R}^n \setminus 0$. Assume $\det \psi_{x\theta} \neq 0$ on a small conic patch. The parameter θ need not be the spatial covector; that covector is $\partial_x\psi$.

To use this initial phase, replace the usual Fourier kernel on the input side as well. We look for terms of the form

$$u_0(t,x) = (2\pi)^{-n} \iint e^{\mathrm{i}(\phi(t,x,\theta)-\psi(y,\theta))} a(t,x,\theta) f(y)\, \mathrm{d}y\, \mathrm{d}\theta. \tag{3.1}$$

We may allow a to depend on y as well. The eikonal equation for ϕ is still (II.2.11), and the transport equations are (II.4.2), (II.4.3), with $L_\pm$ given by (II.4.1). The new initial conditions, replacing (II.2.12), are

$$\phi_\pm|_{t=0} = \psi(x,\theta). \tag{3.2}$$

The method of characteristics applies as before. At $t = 0$ the kernel phase is $\psi(x,\theta) - \psi(y,\theta)$. To match the initial data, we must therefore understand operators of the form

$$Af(x) = (2\pi)^{-n} \iint e^{\mathrm{i}(\psi(x,\theta)-\psi(y,\theta))} a(x,y,\theta) f(y)\, \mathrm{d}y\, \mathrm{d}\theta.$$

Near the diagonal, write

$$\psi(x,\theta) - \psi(y,\theta) = (x-y)\cdot\eta, \qquad \eta = \int_0^1 \partial_x\psi(y + s(x-y),\theta)\, \mathrm{d}s.$$

The change of variables $\theta \mapsto \eta$ is locally invertible by our determinant assumption. Thus A is a pseudodifferential operator there. For an amplitude of order zero, its leading symbol is

$$\frac{a^{(0)}(x,x,\theta)}{|\det \psi_{x\theta}(x,\theta)|}, \qquad \eta = \partial_x\psi(x,\theta).$$

To obtain the identity to leading order, choose $a^{(0)}(x,x,\theta) = |\det \psi_{x\theta}(x,\theta)|$. The lower order terms can be chosen successively to give the identity modulo a smooth kernel. For the full wave equation we use both branches and match u and u_t as in Chapter II, Section 4. The two initial

time derivatives $\partial_t \phi_\pm = \mp b(x, \partial_x \psi)$ are distinct and nonzero, so the two leading trace equations can be solved; the lower order equations are solved in the same way.

For example, let κ be a local change of coordinates and take

$$\psi(x, \theta) = \kappa(x) \cdot \theta.$$

With amplitude $|\det D\kappa(y)|$ and integration over all θ, Fourier inversion in the κ coordinates gives the identity at $t = 0$. With conic cutoffs, it gives the identity microlocally where those cutoffs equal one. This is an explicit choice of initial phase other than $x \cdot \theta$.

The same wave front calculation now gives

$$C^\psi_\pm(t) = \big\{(x, \partial_x \phi_\pm; y, \partial_y \psi)\,;\, \partial_\theta \phi_\pm = \partial_\theta \psi,\ \theta \neq 0\big\}.$$

Initially this is the diagonal on our patch. Along a ray, $\partial_\theta \phi_\pm$ keeps its initial value $\partial_\theta \psi(y, \theta)$, while the initial covector is $\partial_y \psi(y, \theta)$. Consequently, $C^\psi_\pm(t)$ is the same ray-flow graph as before, expressed with different parameters. Changing the initial phase changes the representation and the initial amplitudes; it does not change which input and output covectors are related.

4. Fourier integral operators

The preceding examples suggest that the object to keep is the relation between covectors, while the phase may vary from patch to patch. We now give the general definitions. Let X, Y be smooth manifolds of dimensions n_X, n_Y and fix smooth positive densities to write the kernels. All phase formulas below are in local coordinates.

Nondegenerate phase functions. A real phase function $\Psi(x, y, \theta)$ is smooth on an open conic subset of $X \times Y \times (\mathbf{R}^N \setminus 0)$, has nonzero full differential, and satisfies

$$\Psi(x, y, s\theta) = s\Psi(x, y, \theta), \qquad s > 0.$$

The number N of phase variables need not equal the dimension of X or Y. Put

$$\Sigma_\Psi = \{(x, y, \theta)\,;\ \partial_\theta \Psi = 0\}.$$

The phase is *nondegenerate* if the N differentials

$$\mathrm{d}_{x,y,\theta}(\partial_{\theta_1} \Psi), \ldots, \mathrm{d}_{x,y,\theta}(\partial_{\theta_N} \Psi)$$

are linearly independent on Σ_Ψ. Thus Σ_Ψ is a smooth manifold of dimension $n_X + n_Y$. This condition does not require the Hessian $\Psi_{\theta\theta}$ to be invertible. For the operators considered here, we also assume $\partial_x \Psi \neq 0$ and $\partial_y \Psi \neq 0$ on Σ_Ψ, to keep both covectors away from zero.

Canonical relations. Use coordinates (x, η) on T^*X and (y, ζ) on T^*Y. A *canonical relation* from $T^*Y \setminus 0$ to $T^*X \setminus 0$ is a smooth conic Lagrangian submanifold

$$C \subset (T^*X \setminus 0) \times (T^*Y \setminus 0)$$

for the symplectic form $\omega_X - \omega_Y$, where

$$\omega_X = \sum_{j=1}^{n_X} \mathrm{d}\eta_j \wedge \mathrm{d}x^j, \qquad \omega_Y = \sum_{j=1}^{n_Y} \mathrm{d}\zeta_j \wedge \mathrm{d}y^j.$$

Here *conic* means invariance under simultaneous positive scaling of η and ζ. The word *Lagrangian* means that C has dimension $n_X + n_Y$, half the ambient dimension, and that $\omega_X - \omega_Y$ vanishes

on pairs of tangent vectors to C. For a graph, this says that the corresponding transformation preserves the symplectic form. Such a transformation is called canonical.

A nondegenerate phase parametrizes such a relation by

$$C_\Psi = \big\{(x, \partial_x\Psi; y, -\partial_y\Psi)\,;\ (x,y,\theta) \in \Sigma_\Psi\big\}.$$

Let us verify both the dimension and the Lagrangian condition, as we did for the wave kernel. Define the map

$$j : \Sigma_\Psi \longrightarrow (T^*X \setminus 0) \times (T^*Y \setminus 0), \qquad j(x,y,\theta) = (x, \partial_x\Psi; y, -\partial_y\Psi).$$

The rank condition in the definition of a nondegenerate phase gives

$$\dim \Sigma_\Psi = n_X + n_Y + N - N = n_X + n_Y.$$

We must also check that j does not lose rank. Suppose a tangent vector $(\delta x, \delta y, \delta\theta)$ to Σ_Ψ is mapped to zero by $\mathrm{d}j$. The base components give $\delta x = \delta y = 0$. The covector components, together with tangency to $\partial_\theta\Psi = 0$, then give

$$\begin{pmatrix} \Psi_{x\theta} \\ \Psi_{y\theta} \\ \Psi_{\theta\theta} \end{pmatrix} \delta\theta = 0.$$

This matrix is the transpose of the Jacobian of $(x,y,\theta) \mapsto \partial_\theta\Psi$. It has rank N by nondegeneracy, so $\delta\theta = 0$. Thus $\mathrm{d}j$ is injective: j is an immersion. After restricting to a small patch, its image is a submanifold of dimension $n_X + n_Y$.

For the forms, put $\alpha = \eta \cdot \mathrm{d}x - \zeta \cdot \mathrm{d}y$. On Σ_Ψ the θ-derivatives of Ψ vanish, and the sign in the input covector gives

$$j^*\alpha = \big(\partial_x\Psi \cdot \mathrm{d}x + \partial_y\Psi \cdot \mathrm{d}y\big)\big|_{T\Sigma_\Psi} = \mathrm{d}(\Psi|_{\Sigma_\Psi}).$$

Euler's identity says $\Psi = \theta \cdot \partial_\theta\Psi$, so $\Psi|_{\Sigma_\Psi} = 0$. Consequently,

$$j^*(\omega_X - \omega_Y) = j^*(\mathrm{d}\alpha) = \mathrm{d}(j^*\alpha) = 0.$$

Together with the dimension count, this proves that C_Ψ is Lagrangian. Homogeneity also makes it conic: $\partial_\theta\Psi$ has degree zero, while $\partial_x\Psi$ and $\partial_y\Psi$ have degree one.

Thus the sign reversal on the input covector, the critical-point equations, and the rank condition have separate roles. They give the appropriate symplectic form, its vanishing on the relation, and the required half dimension. Different nondegenerate phases can describe the same relation, including the phases $\phi_\pm - \psi$ considered above.

Fourier integral operators. A continuous linear operator $A : C_0^\infty(Y) \to \mathcal{D}'(X)$ is an FIO of order m associated with C if, in local coordinate and conic patches and modulo a smooth kernel, its distribution kernel is a finite sum of oscillatory integrals

$$K_A(x,y) = (2\pi)^{-N} \int e^{\mathrm{i}\Psi(x,y,\theta)} a(x,y,\theta)\,\mathrm{d}\theta,$$

where each Ψ is a nondegenerate phase parametrizing a piece of C, and a is a symbol of order

$$\mu = m + \frac{n_X + n_Y}{4} - \frac{N}{2}.$$

More explicitly, after the local cutoffs, for every compact $K \subset X \times Y$ and all multi-indices α, β,

$$|\partial_{x,y}^\alpha \partial_\theta^\beta a(x,y,\theta)| \le C_{K,\alpha,\beta}(1 + |\theta|)^{\mu - |\beta|}, \qquad (x,y) \in K.$$

We cut off near $\theta = 0$ and keep the conic support inside the domain of the phase. The integral is understood in the distribution sense: insert $\chi(\varepsilon\theta)$, with $\chi \in C_0^\infty(\mathbf{R}^N)$ equal to one near zero, and let $\varepsilon \to 0$. The result is independent of χ.

In our constructions the symbols are *classical*:

$$a \sim \sum_{j=0}^{\infty} a_{\mu-j}, \qquad a_{\mu-j}(x, y, s\theta) = s^{\mu-j} a_{\mu-j}(x, y, \theta)$$

at large frequencies. The remainder after the first M terms is a symbol of order $\mu - M$. The shift between m and μ makes the operator order independent of the choice and number of phase variables. When $n_X = n_Y = N = n$, as in (1.1), the FIO order is just the amplitude order. A smooth kernel may always be added without changing the FIO class.

The local calculation in Section 1 is the model for the general wave front estimate

$$\mathrm{WF}(Af) \subset C \circ \mathrm{WF}(f), \qquad C \circ \mathrm{WF}(f) = \{(x, \eta)\,;\ (x, \eta; y, \zeta) \in C \text{ for some } (y, \zeta) \in \mathrm{WF}(f)\}.$$

Here $f \in \mathcal{E}'(Y)$, or A is properly supported if we allow arbitrary distribution data. Proper support means that each projection of $\operatorname{supp} K_A$ has compact inverse images of compact sets. The nonzero input and output covector assumptions exclude contributions with a zero covector on one side.

5. Changing phase functions and caustics

For the wave operator, the phase $\phi_\pm - y \cdot \xi$ is nondegenerate because

$$\partial_{y^k} \partial_{\xi_j} (\phi_\pm - y \cdot \xi) = -\delta_{jk}.$$

Thus the general definition includes our original local operator. The same argument applies to $\phi_\pm - \psi$ when $\det \psi_{y\theta} \neq 0$. For fixed t, the branches acting on f_1 in (II.2.5) have order zero, and those acting on f_2 have order -1 because of $|\xi|_g^{-1}$.

What can fail when we increase t? Write the ray flow as

$$\Phi_{\pm t}(y, \xi) = \big(x(t; y, \xi), \eta(t; y, \xi)\big).$$

For each fixed t, its graph is smooth as long as the Hamilton flow is defined. The flow is invertible, with inverse $\Phi_{\mp t}$. The phase $\phi_\pm(t, x, \xi) - y \cdot \xi$, however, uses (x, ξ) as coordinates on that graph. To construct it, we inverted

$$(y, \xi) \longmapsto \big(x(t; y, \xi), \xi\big).$$

This requires $\det D_y x(t; y, \xi) \neq 0$. When that determinant vanishes, these particular coordinates fail. We can still use (y, ξ) to describe the graph, but we may no longer be able to write it with one smooth phase of the form $\phi_\pm(t, x, \xi) - y \cdot \xi$.

This is the distinction already encountered in Chapter II, Section 3. A caustic is a singularity of the relevant projection of the ray geometry. For a particular incoming wave front, its projection to physical space may become singular even when a phase for the propagator is still available. The inward spherical front in the Euclidean example focuses, while $x \cdot \xi \mp t|\xi|$ remains a phase for the propagator for every t. In either situation we must keep the covectors: rays meeting in space can be distinct in the cotangent bundle.

Changing the phase. Let us see how another phase can describe a smooth canonical relation when the old coordinates fail. Reverse the input covector and regard C as a Lagrangian submanifold

Λ of $T^*(X \times Y)$. Put

$$z = (x, y), \qquad p = (\eta, -\zeta).$$

Thus the canonical one-form is $p \cdot \mathrm{d}z = \eta \cdot \mathrm{d}x - \zeta \cdot \mathrm{d}y$. It vanishes on Λ. Indeed, the radial vector field $E = \sum p_j \partial_{p_j}$ is tangent to a conic Λ, and

$$\big(\mathrm{d}(p \cdot \mathrm{d}z)\big)(E, v) = p \cdot \mathrm{d}z(v) = 0, \qquad v \in T\Lambda,$$

by the Lagrangian property.

Locally we can choose some position coordinates and the complementary momentum coordinates as coordinates on Λ. After a choice of base coordinates and a relabeling, write these as (z', p''), with $z = (z', z'')$ and $p = (p', p'')$. This is a consequence of the Lagrangian rank condition: a maximal independent set of position differentials can be completed by the complementary momentum differentials. Shrink to a conic patch where these coordinates work. The radial vector is nonzero, so $p'' \neq 0$ there.

In these coordinates define

$$S(z', p'') = -z''(z', p'') \cdot p''.$$

Since $p' \cdot \mathrm{d}z' + p'' \cdot \mathrm{d}z'' = 0$ on Λ, differentiation gives

$$\mathrm{d}S = p' \cdot \mathrm{d}z' - z'' \cdot \mathrm{d}p'', \qquad p' = \partial_{z'} S, \quad z'' = -\partial_{p''} S.$$

The function S is homogeneous of degree one in p'': scaling the covectors leaves z'' fixed. We can now use the phase

$$\Psi(z, \theta) = S(z', \theta) + z'' \cdot \theta, \qquad \theta = p''.$$

Its critical-point equations and covectors are

$$\partial_\theta \Psi = 0 \iff z'' = -\partial_\theta S, \qquad \partial_z \Psi = (\partial_{z'} S, \theta).$$

They describe precisely the same Λ, and reversing the input sign gives back C. This phase is nondegenerate because

$$\partial_{z''_k} \partial_{\theta_j} \Psi = \delta_{jk}.$$

In particular, $S_{\theta\theta}$ need not be invertible. A singular projection to position space does not prevent us from having a nondegenerate phase.

The variables θ here are momentum coordinates in the directions where we gave up using position coordinates. Integrating in them is a partial inverse Fourier transform. This is the reason for introducing extra phase variables: they let us describe a wave near a caustic without solving for all the momenta, or all the original ray parameters, as functions of position. On another patch we may need another splitting, and hence another phase. This changes the representation of the same canonical relation.

Matching the amplitudes. Keeping the canonical relation is necessary, but it does not by itself make two oscillatory integrals the same operator. We must change their amplitudes as well. For a homogeneous invertible change of phase variables $\theta = \Theta(z, \vartheta)$, the rule is just change of variables:

$$\widetilde{\Psi}(z, \vartheta) = \Psi(z, \Theta(z, \vartheta)), \qquad \widetilde{a}(z, \vartheta) = a(z, \Theta(z, \vartheta)) \, |\det \Theta_\vartheta(z, \vartheta)|.$$

More general changes of phase also involve stationary phase, possibly after introducing additional variables. The local equivalence theorem, see Hörmander [5, Sections 3.1–3.2], says that two nondegenerate phases parametrizing the same piece of Λ give the same class of kernels: an amplitude

for one can be replaced by an amplitude for the other so that the kernels agree modulo a smooth kernel. If the numbers of phase variables are N and $\widetilde{N}$, the amplitude orders change according to

$$\widetilde{\mu} = \mu + \frac{N - \widetilde{N}}{2}.$$

The FIO order m therefore stays the same.

There is a useful point here about caustics. After separating a large frequency $\lambda > 0$, suppose we eliminate k variables v by stationary phase. If v_0 is a critical point of $F(z, v)$ and $H = F_{vv}(z, v_0)$ is invertible, the leading term of the integral is

$$\int e^{\mathrm{i}\lambda F(z,v)} b(z, v, \lambda)\, \mathrm{d}v \;\sim\; \left(\frac{2\pi}{\lambda}\right)^{k/2} e^{\mathrm{i}\lambda F(z,v_0)} \frac{e^{\mathrm{i}\pi \operatorname{sgn} H/4}}{|\det H|^{1/2}}\, b(z, v_0, \lambda).$$

Here the amplitude is localized near v_0; lower order terms are obtained by differentiating b. The signature $\operatorname{sgn} H$ is the number of positive eigenvalues minus the number of negative ones. The determinant and signature factors must be included when matching representations.

If the Hessian used in this reduction becomes singular, the reduced formula is no longer valid. For example, two stationary points can merge at a caustic. We then keep those integration variables and, when needed, use a different phase patch. The full phase can remain nondegenerate in the sense defined above even though this smaller Hessian is singular. Thus the apparent blow-up of a reduced amplitude need not be a failure of the oscillatory integral representing the solution. The signature factors acquired when changing phases are the source of the Maslov phase factors; we do not develop that bookkeeping here.

6. Continuation through a spherical focus

Return to the Euclidean wave equation in $\mathbf{R}^3$ and the inward spherical front from Chapter II, Section 3. We can see the change of phase explicitly. Let $\chi \in C^\infty([0, \infty))$ vanish near zero and equal one for large arguments, and set

$$u(t, x) = (2\pi)^{-3} \int_{\mathbf{R}^3} e^{\mathrm{i}\Psi(t,x,\xi)} \chi(|\xi|)\, \mathrm{d}\xi, \qquad \Psi(t, x, \xi) = x \cdot \xi - (t - 1)|\xi|.$$

The integral is understood as an oscillatory integral. Each exponential solves $(\partial_t^2 - \Delta)u = 0$, so u is an exact distributional solution. It is the kernel of the plus half-wave propagator from time 1 to time t, with the frequency cutoff and with the source point fixed at the origin. We chose the shift by 1 so that the incoming front at $t = 0$ is the unit sphere.

First let us recover the radial phases away from $x = 0$. Write $\xi = \lambda\omega$, $\lambda > 0$, $\omega \in S^2$. Rotating the polar axis to the direction of x, we get

$$\int_{S^2} e^{\mathrm{i}\lambda x \cdot \omega}\, \mathrm{d}\omega = 2\pi \int_{-1}^{1} e^{\mathrm{i}\lambda |x| s}\, \mathrm{d}s = 4\pi\, \frac{\sin(\lambda|x|)}{\lambda|x|}.$$

Consequently, for $x \neq 0$,

$$\begin{aligned} u(t, x) &= \frac{1}{2\pi^2 |x|} \int_0^\infty e^{\mathrm{i}\lambda(1-t)} \lambda \chi(\lambda) \sin(\lambda|x|)\, \mathrm{d}\lambda \\ &= \frac{1}{4\pi^2 \mathrm{i}|x|} \int_0^\infty \lambda\chi(\lambda) \left(e^{\mathrm{i}\lambda(1-t+|x|)} - e^{\mathrm{i}\lambda(1-t-|x|)} \right) \mathrm{d}\lambda. \end{aligned}$$

This identity is exact, not just a stationary phase approximation. It gives the two phases

$$\psi_{\text{in}}(t,x,\lambda) = \lambda(1-t-|x|), \qquad \psi_{\text{out}}(t,x,\lambda) = \lambda(1-t+|x|).$$

For $t<1$, near $|x| = 1-t$, only the incoming term contributes to the wave front set: the outgoing term has nonzero λ-derivative and is smooth there by integration by parts. For $t>1$, the singular front is $|x| = t-1$ and comes from the outgoing term. Both have temporal covector $-\lambda$, so both belong to our plus half-wave branch. Incoming and outgoing here describe motion toward and away from the origin.

What happens at $t=1$, $x=0$? The radial phases are no longer smooth in x, and the amplitudes contain $|x|^{-1}$. Keeping either radial formula alone does not give a continuation through the focus. Instead, return to the phase Ψ with its three integration variables. It is smooth for $\xi \neq 0$, homogeneous of degree one, and

$$\partial_{\xi_j}\Psi = x^j - (t-1)\frac{\xi_j}{|\xi|}, \qquad \partial_{x^k}\partial_{\xi_j}\Psi = \delta_{jk}.$$

Thus it is nondegenerate even at the focus. We have replaced the singular radial description by a regular phase with more variables.

The stationary equations give

$$x = (t-1)\frac{\xi}{|\xi|}, \qquad (\tau,\eta) = \big(-|\xi|,\xi\big).$$

Writing $\xi = \lambda\omega$, the associated Lagrangian is

$$\Lambda = \big\{(t,(t-1)\omega;-\lambda,\lambda\omega)\,;\ \omega\in S^2,\ \lambda>0\big\}.$$

It is smooth through $t=1$: the parameters (t,ω,λ) are recovered from t and η. Its projection to spacetime loses the angular directions at the focus, since all ω give $x=0$ there. The covectors $\eta = \lambda\omega$ still distinguish the rays. Before the focus, η points toward the origin; after it, the same ray and covector point outward.

The amplitude matching is visible as well. For $x\neq 0$, the angular phase $x\cdot\omega$ has critical points $\omega = \pm x/|x|$. Eliminating the two angular variables gives the factor $(\lambda|x|)^{-1}$ and the relative minus sign in the radial formula. At $x=0$ the angular phase is constant, so this reduction is invalid; we keep the angular variables instead. For each fixed λ, the angular integral tends to 4π as $x\to 0$. The relative sign of the outgoing and incoming terms is fixed by this common integral.

This does not make the focused solution smooth. At the focal time,

$$u(1,x) = \delta_0(x) \mod C^\infty,$$

because $\chi(|\xi|)-1$ is smooth and compactly supported. What continues through the caustic is the distribution and its smooth Lagrangian description. Away from the focus, we can reduce it again to the outgoing radial phase.

Finally, allowing the source point to vary replaces x by $x-y$ in Ψ. The corresponding propagator has canonical relation

$$C(t,1) = \big\{(x,\xi;y,\xi)\,;\ x = y + (t-1)\xi/|\xi|,\ \xi\neq 0\big\}.$$

Here it is the spherical front that focuses; the Euclidean propagator phase remains available throughout. The change from the incoming radial phase to Ψ, and then to the outgoing radial phase, gives a concrete instance of continuing the same wave by changing its phase representation.

7. The propagator on the sphere

We now give an example where the phase $\phi_+(t,x,\xi) - y\cdot\xi$ of the propagator itself breaks down. On the unit round sphere with metric g, take $c = 1$ and $g_0 = g$ in the notation of Chapter II. We study

$$B = \sqrt{-\Delta_g}, \qquad U_+(t) = e^{-\mathrm{i}tB},$$

with kernels taken relative to the Riemannian volume $\mathrm{d}\,\mathrm{Vol}_g$. The short-time geometric optics construction is local, so it applies in coordinate patches on the sphere as well. In latitude and longitude coordinates $x = (q,z)$, we have

$$g = \mathrm{d}q^2 + \cos^2 q\,\mathrm{d}z^2, \qquad b(q,z,\eta) = |\eta|_g = \left(\eta_q^2 + \frac{\eta_z^2}{\cos^2 q}\right)^{1/2}.$$

We stay near the equator and use a continuous choice of longitude, so there is no coordinate singularity in the region considered. The plus half-wave follows the Hamilton flow of b.

Take initial points $y = (a,w)$, with a small, and hold the initial covector fixed at $\xi = (0,\kappa)$, $\kappa > 0$. These rays start due east with unit speed. To compute them, use the embedding and eastward unit vector

$$X(q,z) = (\cos q\cos z, \cos q\sin z, \sin q), \qquad E(w) = (-\sin w, \cos w, 0).$$

The great circle through $X(a,w)$ with this initial velocity is

$$X(x(t;a,w)) = \cos t\, X(a,w) + \sin t\, E(w).$$

In particular,

$$\sin q(t) = \sin a\cos t, \qquad \cos q(t)\, e^{\mathrm{i}(z(t)-w)} = \cos a\cos t + \mathrm{i}\sin t.$$

At $a = 0$ the ray is $q(t) = 0$, $z(t) = w + t$. Differentiating the preceding formulas with respect to (a,w), with ξ fixed, gives

$$D_y x(t;0,w,0,\kappa) = \begin{pmatrix} \cos t & 0 \\ 0 & 1 \end{pmatrix}, \qquad \det D_y x = \cos t.$$

Thus at $t_* = \pi/2$ the mixed projection $(y,\xi) \mapsto (x(t;y,\xi),\xi)$ is singular. This is exactly the projection used to construct $\phi_+(t,x,\xi)$ from its initial value $x\cdot\xi$.

We can see the failure without just computing a determinant. At $t = t_*$ all the initial points (a,w), with w fixed and with this same covector, arrive at

$$x = (0, w + \pi/2).$$

Their energy is $b(y,\xi) = \kappa/\cos a$. Since $\dot q(t_*) = -\sin a$ and $\dot z(t_*) = \cos a$, the final covectors are

$$\eta = (-\kappa\tan a, \kappa).$$

The same (x,ξ) therefore corresponds to different η and different y. It cannot parametrize the graph near this ray. A smooth function $\phi_+(t_*,x,\xi)$ cannot produce all these values through $\eta = \partial_x\phi_+$ and $y = \partial_\xi\phi_+$, although the Hamilton flow and its graph are smooth.

There is also an explicit formula for the restriction of the short-time phase to $\xi = (0,\kappa)$. For $0 < t < \pi/2$ and q sufficiently small,

$$\phi_+(t,q,z;0,\kappa) = \kappa\left(z - \arcsin\frac{\sin t}{\cos q}\right).$$

Indeed, ϕ_+ is constant along each ray and equals its initial value κw; the ray formulas give $\cos q \sin(z-w) = \sin t$. We choose the branch $z - w < \pi/2$ before t_*. The formula has initial value κz and satisfies the plus eikonal equation. At the central ray,

$$\partial_q^2 \phi_+(t, 0, z; 0, \kappa) = -\kappa \tan t.$$

This derivative blows up as $t \uparrow t_*$. The propagator phase itself has reached the limit of this representation.

Let us replace it. Near these pairs of points the spherical distance is smooth:

$$d(x, y) = \arccos\big(\sin q \sin a + \cos q \cos a \cos(z - w)\big).$$

Here d is close to $\pi/2$, away from both 0 and the antipodal distance π. Use the phase

$$\Psi(t, x, y, \lambda) = \lambda\big(d(x, y) - t\big), \qquad \lambda > 0.$$

It is homogeneous of degree one in λ, and

$$\partial_\lambda \Psi = d(x, y) - t, \qquad (\tau, \eta; \xi) = \big(-\lambda, \lambda\, \mathrm{d}_x d; -\lambda\, \mathrm{d}_y d\big).$$

For each fixed t near t_*, this is a nondegenerate phase: the differential of $d(x, y) - t$ in (x, y) is nonzero, since $|\mathrm{d}_x d|_g = 1$. The stationary equation $d(x, y) = t$ and the endpoint covectors describe the unit-speed geodesic from y to x. Hence this phase parametrizes the same graph of Φ_t.

Check the covectors at the caustic explicitly. At $x = (0, w + \pi/2)$, $y = (a, w)$, differentiation gives

$$d(x, y) = \pi/2, \qquad \mathrm{d}_x d = (-\sin a, \cos a), \qquad -\mathrm{d}_y d = (0, \cos a).$$

Taking $\lambda = \kappa/\cos a$ recovers

$$\xi = (0, \kappa), \qquad \eta = (-\kappa \tan a, \kappa),$$

exactly as above. The temporal covector is $-\lambda$, so this remains our plus half-wave branch.

With local cutoffs, the propagator kernel near this part of the relation can therefore be written

$$K_+(t, x, y) = \int_0^\infty e^{\mathrm{i}\lambda(d(x,y)-t)} A(t, x, y, \lambda)\, \mathrm{d}\lambda \mod C^\infty.$$

At fixed t this has two-dimensional input and output spaces and one phase variable. The amplitude has order $1/2$ for a propagator of order zero, in agreement with the order convention above. Its coefficients remain smooth through t_*. For example, with the kernel taken relative to Riemannian volume, a leading coefficient depending only on d satisfies

$$2a_0'(d) + (\cot d) a_0(d) = 0, \qquad a_0(d) = c_0 (\sin d)^{-1/2},$$

because $\Delta_g d = \cot d$ on the round two-sphere. The constant c_0 is fixed by matching the preceding propagator representation. This coefficient is regular at $d = \pi/2$, and the successive transport equations are regular there as well.

The failure here occurs in the coordinates of the propagator's canonical relation, before applying it to any chosen datum. Passing from $\phi_+(t, x, \xi) - y \cdot \xi$ to $\lambda(d(x, y) - t)$ lets us continue the same kernel across $t = \pi/2$. The caustic is that of the mixed projection (x, ξ); the distance phase works because the endpoints are not yet conjugate. At the antipodal distance π, that phase would in turn require another representation.

7.1. Continuation through the antipode. Let us now reach $t = \pi$. All unit-speed geodesics from y meet at its antipode $\alpha(y) = -y$. The function $d(x, y)$ is not smooth there, and the coefficient $(\sin d)^{-1/2}$ in the preceding representation is singular. We need to keep the direction of the ray as an integration variable.

First identify the canonical relation at this time. Regard the sphere as a subset of $\mathbf{R}^3$. If $v = \xi^\sharp/|\xi|_g$ is the initial unit tangent vector, then

$$x(t) = \cos t\, y + \sin t\, v, \qquad \dot{x}(t) = -\sin t\, y + \cos t\, v.$$

Here $\xi^\sharp$ is the vector obtained from the covector ξ using the metric. At $t = \pi$, both the position and the tangent vector change sign. Consequently,

$$\Phi_\pi(y, \xi) = \big(\alpha(y), (D\alpha_y)^{-T}\xi\big) =: \widehat{\alpha}(y, \xi), \qquad \Phi_{\pi+s} = \Phi_s \circ \widehat{\alpha}.$$

The inverse transpose is the usual transformation rule for covectors. The graph is smooth at $t = \pi$, although its projection with y fixed collapses all directions to $\alpha(y)$.

Put $s = t - \pi$ and $w = \alpha(y)$. Near a fixed antipodal pair use the vector

$$V(w, x) = \exp_w^{-1}(x) \in T_w S^2.$$

It is smooth when x is near w, including $x = w$. On the sphere it is explicit. With $\delta = d(x, w)$ and with x, w regarded as vectors in $\mathbf{R}^3$,

$$V(w, x) = \frac{\delta}{\sin \delta}\big(x - \cos \delta\, w\big), \qquad V(w, w) = 0.$$

The ratio has expansion $1 + \delta^2/6 + \cdots$, so this formula extends smoothly to the diagonal even though the scalar distance δ does not.

Choose a smooth orthonormal frame near w and express V in that frame. Let $\theta \in \mathbf{R}^2 \setminus 0$ denote covector components in its dual frame. We propose the phase

$$\Psi_\pi(t, x, y, \theta) = V(\alpha(y), x) \cdot \theta - (t - \pi)|\theta|.$$

This is the short-time exponential-map phase in Zelditch [13, Section 11.1], with base point $\alpha(y)$ and time $s = t - \pi$. It is homogeneous of degree one in θ and smooth on this patch for $|t - \pi|$ small. Its stationary equations are

$$\partial_\theta \Psi_\pi = 0 \quad \Longleftrightarrow \quad V(w, x) = s\, \frac{\theta}{|\theta|}.$$

Thus x is reached by following the short geodesic from w for the signed time s, with initial covector θ. Both $s < 0$ and $s > 0$ are allowed.

The nondegeneracy check is particularly simple. In these coordinates,

$$D_x(\partial_\theta \Psi_\pi) = D_x V(w, x),$$

which is invertible near $x = w$. Hence the two differentials $\mathrm{d}(\partial_{\theta_j} \Psi_\pi)$ are independent, also at $s = 0$. At that time the stationary set is just $x = \alpha(y)$, with all nonzero θ retained.

Let us check the endpoint covectors, including the input sign. On the stationary set, let $\mathcal{P}_s\theta$ denote parallel transport of θ from w to x along the short geodesic. The first variation formula gives

$$\eta = \partial_x \Psi_\pi = \mathcal{P}_s\theta, \qquad \xi = -\partial_y \Psi_\pi = (D\alpha_y)^T\theta, \qquad \tau = \partial_t \Psi_\pi = -|\theta|.$$

For $s \neq 0$ these identities follow by differentiating $\frac{1}{2}d(w,x)^2 = \frac{1}{2}|V(w,x)|^2$ and using $V = s\theta^\sharp/|\theta|$; they extend to $s = 0$ by continuity. Terms from differentiating the chosen frame cancel on the stationary set. Therefore

$$(w,\theta) = \widehat{\alpha}(y,\xi), \qquad (x,\eta) = \Phi_s(w,\theta) = \Phi_{\pi+s}(y,\xi).$$

This proves that Ψ_π parametrizes the required propagator relation through the antipodal caustic. Also $|\eta|_g = |\theta| = |\xi|_g$, and $\tau = -|\eta|_g$, as required for the plus branch. This is a general kernel phase; its characteristic condition is imposed on the stationary set.

We can compare it directly with the distance phases on the two sides. For $\delta > 0$, write $\theta = \lambda\omega$, $\lambda > 0$, $|\omega| = 1$. The angular critical points are $\omega = \pm V/|V|$, giving the reduced phases

$$\lambda(\delta - s), \qquad \lambda(-\delta - s).$$

Since $\delta = \pi - d(x,y)$, these become

$$\lambda\big(2\pi - d(x,y) - t\big), \qquad \lambda\big(d(x,y) - t\big).$$

Before the caustic, $s < 0$, the second phase has stationary radial variable, with $\delta = -s$. After it, $s > 0$, the first does, with $\delta = s$. The ray has then traveled a length $t > \pi$, while the shortest distance between its endpoints is $2\pi - t$. At $s = 0$, $\delta = 0$, the angular reduction fails, but the full phase Ψ_π remains nondegenerate. It records how each direction continues past the antipode.

With local cutoffs, the resulting representation is

$$K_+(t,x,y) = (2\pi)^{-2} \int_{\mathbf{R}^2} e^{\mathrm{i}\Psi_\pi(t,x,y,\theta)} a_\pi(t,x,y,\theta)\,\mathrm{d}\theta \mod C^\infty,$$

where a_π is a classical symbol of order zero with coefficients smooth through $t = \pi$. The change from one to two phase variables changes the amplitude order from $1/2$ to zero. The singular factor in the reduced distance amplitude is absent from this representation.

The amplitude must also carry the focusing phase. We can determine it on the round two-sphere without developing the general Maslov theory. To compare B with an operator whose spectrum is simpler, put

$$\widetilde{B} = \sqrt{-\Delta_g + \tfrac{1}{4}}, \qquad \mathcal{A}f(x) = f(\alpha(x)).$$

A spherical harmonic of degree ℓ has $-\Delta_g$ eigenvalue $\ell(\ell+1)$; see Zelditch [13, Sections 4.1–4.2]. Its antipodal parity is $(-1)^\ell$; see NIST DLMF [3, Eq. (14.30.7)]. Thus $\widetilde{B}$ has eigenvalue $\ell + \frac{1}{2}$, and

$$e^{-\mathrm{i}\pi\widetilde{B}} = -\mathrm{i}\,\mathcal{A}.$$

The operators B and $\widetilde{B}$ commute. Set

$$Q = \widetilde{B} - B = \tfrac{1}{4}(\widetilde{B} + B)^{-1}, \qquad R = e^{\mathrm{i}\pi Q}.$$

By the square-root calculus discussed in Chapter II, Section 10, Q is of order -1 and R is an elliptic ΨDO of order zero with $R - \mathrm{Id}$ of order -1. We obtain the exact identities

$$U_+(\pi) = -\mathrm{i}\,\mathcal{A}R, \qquad U_+(\pi + s) = -\mathrm{i}\,U_+(s)\mathcal{A}R.$$

This also constructs the continuation from a short-time propagator. The factor R has principal symbol 1, changes the lower order amplitudes, and keeps the same canonical relation. For Riemannian volume and the orthonormal frame used above, the leading amplitude at $t = \pi$, $x = \alpha(y)$ is $-\mathrm{i}$. Let us check how this value matches the amplitudes on both sides of the caustic. In the calculations below the local cutoffs are equal to one near the ray under consideration.

The leading amplitude. First consider the short-time propagator from w to x, with phase $V(w,x)\cdot\theta - s|\theta|$. Write $\delta = |V(w,x)|$. In geodesic polar coordinates about w, the area element on the unit two-sphere and the Euclidean area element in T_wS^2 are

$$\mathrm{d}\,\mathrm{Vol}_g(x) = \sin\delta\,\mathrm{d}\delta\,\mathrm{d}\omega, \qquad \mathrm{d}V = \delta\,\mathrm{d}\delta\,\mathrm{d}\omega.$$

Thus the volume Jacobian of the exponential map is $J(\delta) = \sin\delta/\delta$. We can also obtain its contribution to the amplitude directly from the transport equation. By rotational symmetry the leading coefficient can be chosen to depend only on δ; denote it by $\beta(\delta)$. Angular stationary phase contributes $\delta^{-1/2}$, so the leading distance coefficient is a constant times $\beta(\delta)\delta^{-1/2}$. Substituting this into the radial transport equation above gives

$$2\beta'(\delta) + \left(\cot\delta - \frac{1}{\delta}\right)\beta(\delta) = 0, \qquad \beta(0) = 1.$$

The initial value is fixed by the identity kernel at $s = 0$. Integration gives

$$\beta(\delta) = J(\delta)^{-1/2} = \left(\frac{\delta}{\sin\delta}\right)^{1/2} = 1 + \frac{\delta^2}{12} + O(\delta^4).$$

The same real factor occurs for negative s, by taking the adjoint of the short-time propagator. Although δ itself is not smooth at $x = w$, this even function of δ is smooth there.

Since the antipodal map preserves Riemannian volume and R has principal symbol 1, the leading amplitude in the phase Ψ_π is, on its stationary set,

$$a_{\pi,0}(t,x,y,\theta) = -\mathrm{i}\left(\frac{\delta}{\sin\delta}\right)^{1/2} = -\mathrm{i}\left(\frac{s}{\sin s}\right)^{1/2}, \qquad \delta = |s|.$$

The ratio at $s = 0$ is understood by continuity. This gives $a_{\pi,0} = -\mathrm{i}$ at the antipode, with no blow-up. We may use $-\mathrm{i}\beta(\delta)$ as a smooth extension of this leading amplitude off the stationary set. A different extension changes the lower order terms, which we match below.

Matching the two distance phases. For $\delta > 0$, choose angular coordinates in the tangent plane so that

$$V = \delta(1,0), \qquad \theta = \lambda(\cos v, \sin v), \qquad \Psi_\pi = \lambda(\delta\cos v - s), \qquad \mathrm{d}\theta = \lambda\,\mathrm{d}\lambda\,\mathrm{d}v.$$

The angular critical points are $v = 0, \pi$, with second derivatives $-\delta, +\delta$, respectively. Consequently,

$$\int_0^{2\pi} e^{\mathrm{i}\lambda\delta\cos v}\,\mathrm{d}v \sim \left(\frac{2\pi}{\lambda\delta}\right)^{1/2}\left(e^{\mathrm{i}\lambda\delta - \mathrm{i}\pi/4} + e^{-\mathrm{i}\lambda\delta + \mathrm{i}\pi/4}\right).$$

With angular cutoffs we retain the corresponding critical point. Combining this factor with $-\mathrm{i}\beta(\delta)$, the polar Jacobian λ, and the normalization $(2\pi)^{-2}$ gives the leading radial coefficients. Put $d = d(x,y) = \pi - \delta$. Before the caustic the relevant critical point is $v = \pi$, with phase $\lambda(d - t)$; after it, the relevant point is $v = 0$, with phase $\lambda(2\pi - d - t)$. Their amplitudes are

$$A_{\mathrm{before}}(t,x,y,\lambda) = \frac{e^{-\mathrm{i}\pi/4}}{(2\pi)^{3/2}}\frac{\lambda^{1/2}}{\sqrt{\sin d}} + O(\lambda^{-1/2}),$$

$$A_{\mathrm{after}}(t,x,y,\lambda) = \frac{e^{-3\mathrm{i}\pi/4}}{(2\pi)^{3/2}}\frac{\lambda^{1/2}}{\sqrt{\sin d}} + O(\lambda^{-1/2}).$$

Here the remainders are symbols of order $-1/2$ on patches with $\delta > 0$. We used $\sin\delta = \sin d$, $-\mathrm{i}e^{\mathrm{i}\pi/4} = e^{-\mathrm{i}\pi/4}$, and $-\mathrm{i}e^{-\mathrm{i}\pi/4} = e^{-3\mathrm{i}\pi/4}$. This fixes the constant in the preceding distance representation:

$$c_0 = (2\pi)^{-3/2} e^{-\mathrm{i}\pi/4}.$$

The leading radial coefficients on the two sides differ by the factor $-\mathrm{i}$. This is the focusing phase for our choice $U_+(t) = e^{-\mathrm{i}tB}$. These reduced expansions are not uniform as $\delta \to 0$: the angular Hessian then vanishes. The full phase Ψ_π and its amplitude remain smooth in their parameters through the caustic, while the kernel remains a singular distribution.

Matching to all orders. Matching the leading coefficient leaves lower order singular terms. To obtain equality modulo a smooth kernel, take a short-time parametrix $V_+(s)$ for $U_+(s)$, constructed by solving the eikonal and all the transport equations in a regular short-time phase. Its error $V_+(s) - U_+(s)$ has a smooth kernel, smoothly depending on s. The exact identity above shows that

$$W_\pi(t) := -\mathrm{i}\, V_+(t-\pi)\mathcal{A}R, \qquad W_\pi(t) - U_+(t) \quad \text{has a smooth kernel.}$$

We now rewrite the localized kernel of W_π in the nondegenerate phase Ψ_π. Composition and change of phase give a full classical amplitude; see Hörmander [5, Theorems 3.2.1 and 4.2.2]. In this step we use the full symbol of R, including the derivative terms in the composition formula. The phase Ψ_π only has to satisfy the characteristic equation on its stationary set; we obtain its amplitude by rewriting the short-time construction, rather than imposing the earlier pointwise eikonal equation away from that set.

More explicitly, write

$$a_\pi \sim a_{\pi,0} + a_{\pi,-1} + a_{\pi,-2} + \cdots, \qquad a_{\pi,-j}(t,x,y,\lambda\theta) = \lambda^{-j} a_{\pi,-j}(t,x,y,\theta).$$

After choosing the leading term just computed, the difference from W_π is an FIO of order -1 at fixed t. Express its leading symbol in the same phase and choose $a_{\pi,-1}$ to remove it. Repeat: after N terms the remaining FIO has order $-N$. The phase is nondegenerate also at $t = \pi$, so these corrections can be chosen smoothly in t through that time. Their lower order values include the corrections from R, from composition, and from changing the phase; they are not determined by the value $a_{\pi,0} = -\mathrm{i}$ alone.

All symbol estimates here are locally uniform in t near π, including after any number of t-derivatives. Asymptotic summation with t as a smooth parameter gives a symbol a_π such that

$$a_\pi - \sum_{j=0}^{N-1} a_{\pi,-j} \in S^{-N} \quad \text{for every } N.$$

Differentiating a remainder integral k times in (t,x,y) raises its symbol order by at most k. Since the matching is to arbitrary order, the resulting kernel agrees with W_π, and hence with $U_+(t)$, modulo a kernel smooth jointly in (t,x,y) across $t = \pi$ on the patch. The series need not converge. Away from the antipode, applying the full angular stationary-phase expansion gives the distance amplitudes on each side to all orders. At the antipode we retain the angular variable. This is how the same half-wave propagator, with its matched amplitudes, continues through the caustic.

8. Global operators and continuation of the propagator

Let us return to a general canonical relation C. The examples illustrate how compatible local representations give one operator. Cover C by the phase patches of Section 5 and use a microlocal

partition of unity. For an existing operator, we localize its kernel in position and covector variables. In each patch write

$$K_j(x,y) = (2\pi)^{-N_j} \int e^{\mathrm{i}\Psi_j(x,y,\theta_j)} a_j(x,y,\theta_j)\,\mathrm{d}\theta_j.$$

The global kernel is the sum of these localized kernels, modulo a smooth kernel. The sum is locally finite; on a compact part of the base and of the covector directions, finitely many patches suffice. The corresponding pieces of the canonical relation all belong to the same geometric C.

If instead we construct the operator patch by patch, the local representations of the kernel must agree microlocally on overlaps, modulo smooth terms, using the amplitude transformations of Section 5. Arbitrary amplitudes on the same relation will not do. The partition of unity is applied to these compatible representations, so it does not count the overlap twice. A single global phase, or a single scalar amplitude in fixed coordinates, is not required. The global object is the operator together with its canonical relation. For the global construction and the symbol transformation law, see Hörmander [5, Section 3.2].

Continuing the wave propagator. For the wave equation there is a direct way to produce these compatible local kernels. Divide a finite time interval into short pieces and compose the short-time propagators, with local cutoffs along the rays. For two such pieces of lengths s and t, the combined kernel phase on one branch is

$$\phi_\pm(s,x,\theta) - z\cdot\theta + \phi_\pm(t,z,\xi) - y\cdot\xi,$$

where z is the intermediate position. Stationarity in (θ, z, ξ) gives

$$z = \partial_\theta\phi_\pm(s,x,\theta), \qquad \theta = \partial_z\phi_\pm(t,z,\xi), \qquad y = \partial_\xi\phi_\pm(t,z,\xi).$$

Thus the output covector of the first step is the input covector of the next:

$$(z,\theta) = \Phi_{\pm t}(y,\xi), \qquad (x,\eta) = \Phi_{\pm s}(z,\theta) = \Phi_{\pm(s+t)}(y,\xi).$$

For graphs of canonical transformations this matching is transverse, meaning that the linearized matching equations have full rank. Stationary phase in suitable local variables then gives an FIO associated with the composed graph; see Hörmander [5, Section 4.2]. We need not reduce its phase back to the special form $\phi_\pm(s+t,x,\xi) - y\cdot\xi$. Near a caustic we use a phase that is regular there, as above.

Repeating this construction continues the parametrix for every finite time for which the flow under consideration is defined. The half-wave group law in Chapter II, Section 10 identifies it, modulo smoothing terms, with the exact propagator $U_\pm(t)$ under the assumptions made there. Its canonical relation is still the graph of $\Phi_{\pm t}$. Changing phases therefore lets the oscillatory description pass through a caustic while following the same rays and preserving the two half-wave signs. This gives continuation as a distribution; it does not assert that a singular wave becomes smooth at its focus.

Gaussian beams

Most of this chapter is AI generated after a series of prompts. Read at your own peril!

We return to the two parameter expansion of Chapter II, Section 8. The critical scale is $\varepsilon = \lambda^{-1/2}$. Then we have

$$u|_{t=0} = \chi(\lambda^{1/2}x)e^{-\mathrm{i}\lambda x\cdot\omega}. \tag{0.1}$$

At this scale the bounds obtained there have the same order for every k. Recall the coherent state in Chapter I. Its Gaussian factor can be included in the exponential by giving the phase a positive imaginary quadratic part. For a Gaussian beam we determine the phase and amplitude along a ray so that the eikonal and transport errors vanish there to the required order. The imaginary part gives Gaussian decay away from the ray.

Chapter III constructed the half-wave propagators as FIOs. Here we follow a family of initial packets of width $O(\lambda^{-1/2})$, with a central spatial covector of size λ. On either half-wave branch the packet center follows the Hamilton flow describing the canonical relation in Chapter III, Section 2. The quadratic part of the complex phase describes localization about this ray. A positive definite initial imaginary Hessian, propagated by the Riccati equation below, keeps the beam construction regular along a fixed ray segment, even through a caustic of the real ray coordinates discussed in Chapter III, Section 5. We construct the phase and amplitudes directly from the eikonal and transport equations along the ray. Each wave is smooth for fixed λ, so its ordinary wave front set is empty. The concentration is seen as $\lambda \to \infty$; the ordinary wave front estimate alone does not measure it.

1. The Euclidean case

For the construction of the phase and amplitudes below we follow Section I of Ralston's notes *Gaussian Beams* [12], with the signs and normalizations used here. For the Euclidean wave equation

$$u_{tt} - \Delta u = 0, \tag{1.1}$$

the time Fourier transform with kernel $e^{\mathrm{i}\lambda t}$ gives the Helmholtz equation away from sources. Let us look at a beam for this stationary equation.

1.1. The Helmholtz equation. We use the time Fourier convention and outgoing Green's function fixed in Chapter II, Section 11, with time factor $e^{-\mathrm{i}\lambda t}$. For this example with an explicit Euclidean source, we follow Liu, Ralston, Runborg, and Tanushev [8, Section 5].

We work in $\mathbf{R}^3$ here for simplicity. Let us say we want to solve

$$(-\Delta - \lambda^2)u = -2\mathrm{i}\lambda e^{-\lambda|x'|^2/2}\delta(x^3), \tag{1.2}$$

with $x = (x', x^3)$. The source is a delta function on the plane $x^3 = 0$ with a Gaussian profile in the variables x' transverse to the beam axis. The factor $-2\mathrm{i}\lambda$ is chosen for convenience.

The outgoing solution is given by the outgoing Green's function (operator), which is a convolution with $e^{\mathrm{i}\lambda|x|}/(4\pi|x|)$ applied to the source

$$u(x) = \frac{-2\mathrm{i}\lambda}{4\pi}\int_{\mathbf{R}^2} \frac{e^{\mathrm{i}\lambda|x-(y',0)|-\lambda|y'|^2/2}}{|x-(y',0)|}\mathrm{d}y'. \tag{1.3}$$

We apply the stationary phase to this integral at $x' = 0$ (the x^3 axis), where most of the energy is concentrated. Assume $x^3 \neq 0$. The phase is

$$\psi = |x-(y',0)| + \mathrm{i}|y'|^2/2$$

and is complex, depending also on the parameter x. This is covered by [6, Theorem 7.7.12]. We have

$$\psi_{y'} = \frac{y'-x'}{|x-(y',0)|} + \mathrm{i}y'.$$

It vanishes only when $x' = y' = 0$, and ψ is real there. For the Hessian, we have

$$\psi_{y'y'}\big|_{x'=y'=0} = (|x^3|^{-1} + \mathrm{i})I_{2\times 2}, \quad \det \psi_{y'y'}\big|_{x'=y'=0} = (|x^3|^{-1}+\mathrm{i})^2.$$

Applying [6, Theorem 7.7.12] with the correct choice of the sign of the square root of the determinant, we get

$$u(0,x^3) = \frac{2\pi}{\lambda}\frac{1}{1-\mathrm{i}/|x^3|}\frac{-2\mathrm{i}\lambda}{4\pi}\frac{e^{\mathrm{i}\lambda|x^3|}}{|x^3|} + O(\lambda^{-1}) = \frac{e^{\mathrm{i}\lambda|x^3|}}{1+\mathrm{i}|x^3|} + O(\lambda^{-1}). \tag{1.4}$$

Here $x^3 \neq 0$ is fixed. The factor $1-\mathrm{i}/|x^3|$ has positive real part and gives the square root with the Gaussian integral convention.

The transverse profile. Let us see what happens near the axis. Work first in $x^3 = s > 0$, with s in a compact subinterval of $(0,\infty)$. Put $x' = \lambda^{-1/2}z$ and $y' = \lambda^{-1/2}\eta$ in (1.3). Expanding the distance and the denominator gives, for bounded z,

$$u(\lambda^{-1/2}z, s) = \frac{-\mathrm{i}e^{\mathrm{i}\lambda s}}{2\pi s}\int_{\mathbf{R}^2} e^{\mathrm{i}|z-\eta|^2/(2s)-|\eta|^2/2}\mathrm{d}\eta + O(\lambda^{-1}).$$

The Gaussian in η controls the tails of the integral. Completing the square, with the same choice of square root as in (1.4), we get

$$u(\lambda^{-1/2}z, s) = \frac{e^{\mathrm{i}\lambda s}}{1+\mathrm{i}s}\exp\left(-\frac{|z|^2}{2(1+\mathrm{i}s)}\right) + O(\lambda^{-1}).$$

Thus the leading beam near the positive axis is

$$u_{\mathrm{GB}}(x', s) = \frac{1}{1+\mathrm{i}s}\exp\left(\mathrm{i}\lambda s - \frac{\lambda|x'|^2}{2(1+\mathrm{i}s)}\right).$$

Its restriction to the axis agrees with (1.4). Moreover,

$$|u_{\mathrm{GB}}(x',s)| = \frac{1}{\sqrt{1+s^2}} \exp\left(-\frac{\lambda|x'|^2}{2(1+s^2)}\right).$$

The transverse width is of order $\lambda^{-1/2}\sqrt{1+s^2}$. It stays of order $\lambda^{-1/2}$ on a fixed bounded interval of s, but it increases as the beam travels. For $x^3 < 0$ the same formula holds with $s = -x^3$. These are the two outgoing beams produced by the source plane. With the time factor $e^{-\mathrm{i}\lambda t}$, the axial phases are $\lambda(s-t)$, in agreement with our outgoing convention. The approximation above is uniform for s bounded away from zero and $|x'| \le C\lambda^{-1/2}$; it is not an estimate at the source plane.

The paraxial approximation. How does this compare with Gaussian beams in optics? There one usually starts with the paraxial approximation to the Helmholtz equation; see, for example, [7, Section 2.4]. Write

$$u(x',s) = e^{\mathrm{i}\lambda s} B(x',s).$$

Away from the source, the full Helmholtz equation becomes

$$\partial_s^2 B + 2\mathrm{i}\lambda \partial_s B + \Delta_{x'} B = 0.$$

If the envelope varies slowly in s compared with the factor $e^{\mathrm{i}\lambda s}$, we neglect $\partial_s^2 B$. This gives the *paraxial equation*

$$2\mathrm{i}\lambda \partial_s B + \Delta_{x'} B = 0.$$

This is precisely the free Schrödinger equation in the transverse variables:

$$\mathrm{i}\partial_s B = -\frac{1}{2\lambda}\Delta_{x'} B.$$

The propagation distance s plays the role of the Schrödinger time. The physical time dependence of the monochromatic wave is still $e^{-\mathrm{i}\lambda t}$. Equivalently, setting $h = \lambda^{-1}$ gives the semiclassical form

$$\mathrm{i}h\partial_s B = -\frac{h^2}{2}\Delta_{x'} B.$$

Taking the Fourier transform in x', we get

$$\widehat{B}(s,\xi') = e^{-\mathrm{i}s|\xi'|^2/(2\lambda)} \widehat{B}(0,\xi'), \qquad B(s) = e^{\mathrm{i}s\Delta_{x'}/(2\lambda)} B(0).$$

In particular, Gaussian initial data stay Gaussian. For $B(x',0) = e^{-\lambda|x'|^2/2}$, evaluating the inverse Fourier transform gives exactly the envelope of our leading beam:

$$B(x',s) = \frac{1}{1+\mathrm{i}s} \exp\left(-\frac{\lambda|x'|^2}{2(1+\mathrm{i}s)}\right).$$

Thus diffraction in this approximation is the spreading of a free Schrödinger Gaussian packet: its transverse width grows by the factor $\sqrt{1+s^2}$. The Gaussian integration also produces the complex prefactor $(1+\mathrm{i}s)^{-1}$. Its absolute value gives the decrease of the amplitude on the axis, while its argument $-\arctan s$ gives the Gouy phase shift described below. Both effects are already contained in the free Schrödinger evolution.

This solves the reduced equation exactly. The approximation was made when we discarded $\partial_s^2 B$ in passing from Helmholtz to Schrödinger. The residual for the full Helmholtz equation is

$$(-\Delta - \lambda^2) u_{\mathrm{GB}} = -e^{\mathrm{i}\lambda s}\partial_s^2 B.$$

In the region $|x'| \le C\lambda^{-1/2}$, with s in a fixed bounded interval, this residual is $O(1)$. The two terms retained in the paraxial equation have scale λ, so the omitted term is smaller by a factor

of order λ^{-1}. This is a relative comparison of the terms in the envelope equation; the absolute residual need not tend to zero. The eikonal and transport construction below supplies corrections for the full equation.

Beam width, curvature, and Gouy phase. Let us separate the absolute value and phase of the same formula. Using $(1+\mathrm{i}s)^{-1} = (1+s^2)^{-1/2}e^{-\mathrm{i}\arctan s}$, we can write

$$u_{\mathrm{GB}}(x',s) = \frac{w_0}{w(s)}\exp\left(-\frac{|x'|^2}{w(s)^2}\right)\exp\left(\mathrm{i}\left[\lambda s + \frac{\lambda|x'|^2}{2R(s)} - \psi(s)\right]\right),$$

where, in our normalization,

$$w_0 = \sqrt{\frac{2}{\lambda}}, \qquad w(s) = w_0\sqrt{1+s^2}, \qquad \frac{1}{R(s)} = \frac{s}{1+s^2}, \qquad \psi(s) = \arctan s.$$

The quantity $w(s)$ is the beam radius at which the field amplitude is e^{-1} times its value on the axis; the squared amplitude is then reduced by e^{-2}. Its minimum w_0 is the *waist radius*. The quadratic transverse phase has coefficient $\lambda/(2R(s))$, and $R(s)$ is the paraxial radius of curvature of the wavefront. At the waist, $1/R(0) = 0$, so the wavefront is flat to this order. The additional axial phase $-\psi(s)$ is the *Gouy phase shift*. Here $\psi(s)$ is positive for $s > 0$, but its contribution to the phase is negative with our time factor $e^{-\mathrm{i}\lambda t}$. Using $e^{\mathrm{i}\lambda t}$ instead conjugates the spatial field and reverses these phase signs.

We have used complex notation. The physical field is the real part:

$$\begin{aligned} u_{\mathrm{phys}}(t,x',s) &= \Re\left(e^{-\mathrm{i}\lambda t}u_{\mathrm{GB}}(x',s)\right) \\ &= \frac{w_0}{w(s)}\exp\left(-\frac{|x'|^2}{w(s)^2}\right)\cos\left(\lambda(s-t) + \frac{\lambda|x'|^2}{2R(s)} - \psi(s)\right). \end{aligned}$$

The Gaussian envelope, curvature, and Gouy shift therefore appear in a real oscillating field. Since the wave operator has real coefficients, taking the real part preserves the equation, and for an approximate solution it preserves the residual bounds.

When discussing intensity, we average over the rapid oscillations in physical time. For the squared field, the average over one period is

$$\begin{aligned} \left\langle u_{\mathrm{phys}}^2\right\rangle_t &:= \frac{\lambda}{2\pi}\int_0^{2\pi/\lambda} u_{\mathrm{phys}}(t,x',s)^2\,\mathrm{d}t \\ &= \frac{1}{2}|u_{\mathrm{GB}}(x',s)|^2 = \frac{1}{2}|B(x',s)|^2. \end{aligned}$$

For the scalar wave equation, we can compute the energy flux explicitly. Write $v = u_{\mathrm{phys}}$ and $U = u_{\mathrm{GB}}$. With $c = 1$, use the energy density and flux

$$e = \frac{1}{2}\left(v_t^2 + |\nabla v|^2\right), \qquad J = -v_t\nabla v.$$

Then $\partial_t e + \operatorname{div} J = v_t(v_{tt} - \Delta v)$, so the energy is locally conserved for an exact solution. This is the normalization used in (V.5.2). The energy E_U defined in Chapter II, Section 1 omits the factor $1/2$; its corresponding flux is $2J$. This choice of energy normalization is separate from the factor $1/2$ obtained by averaging the squared field above.

For $v = \Re(e^{-\mathrm{i}\lambda t}U)$, averaging the product of the two real derivatives gives the exact identity

$$\langle J\rangle_t = -\frac{1}{2}\Re\left((-\mathrm{i}\lambda U)\overline{\nabla U}\right) = \frac{\lambda}{2}\Im\left(\overline{U}\nabla U\right).$$

In particular, since $U = e^{\mathrm{i}\lambda s}B$, the component in the direction of increasing s is

$$\langle J_s\rangle_t = \frac{\lambda^2}{2}|B|^2 + \frac{\lambda}{2}\Im\left(\overline{B}\,\partial_s B\right).$$

These averaging identities apply to our approximate beam as well; they do not require it to solve the full wave equation exactly. For our Gaussian, $\partial_s B/B = O(1)$ when $|x'| \le C\lambda^{-1/2}$ and s stays in a fixed bounded interval. Therefore,

$$\langle J_s\rangle_t = \frac{\lambda^2}{2}|B|^2\left(1 + O(\lambda^{-1})\right).$$

Thus the leading scalar axial flux is λ^2 times the mean squared field. The Gaussian intensity profile describes this leading flux after the frequency factor is included; the envelope derivative supplies the lower order correction. For an optical electric-field amplitude in a homogeneous lossless medium, the leading paraxial intensity is also proportional to the mean squared field, with a factor determined by the medium; see [7, Section 2.1.3]. These time-averaged profiles are stationary for a monochromatic beam, although the real field oscillates. Their dependence on s describes how they change along the beam axis.

The normalization has fixed one length scale. For an arbitrary waist radius w_0, the same paraxial calculation gives

$$s_R = \frac{\lambda w_0^2}{2}, \qquad w(s) = w_0\sqrt{1 + (s/s_R)^2}, \qquad \frac{1}{R(s)} = \frac{s}{s^2 + s_R^2}, \qquad \psi(s) = \arctan(s/s_R).$$

The *Rayleigh range* s_R is the distance at which the radius has increased by a factor $\sqrt{2}$. Our choice $w_0 = \sqrt{2/\lambda}$ gives $s_R = 1$ in the coordinates used above. Thus, at this localization scale, diffraction changes the width over propagation distances of order one. In optics the wave number is usually denoted by k, so $k = \lambda$ in our notation; the wavelength here is $2\pi/\lambda$. The small-angle condition for the paraxial approximation is $\lambda w_0 \gg 1$, which is satisfied by our choice as $\lambda \to \infty$.

We can also see the small parameter directly. Set $X = x'/w_0$, $S = s/s_R$, and $\mathcal{B}(X, S) = B(w_0 X, s_R S)$. The full envelope equation becomes

$$\frac{1}{(\lambda w_0)^2}\partial_S^2\mathcal{B} + \mathrm{i}\partial_S\mathcal{B} + \frac{1}{4}\Delta_X\mathcal{B} = 0.$$

For the Gaussian envelope on bounded sets of (X, S), the derivatives in this equation have scale one. Thus the relative size of the omitted term on the Rayleigh scale is $O((\lambda w_0)^{-2})$, which is $O(\lambda^{-1})$ for our choice of w_0.

There are two different uses of the axial coordinate here. For the source problem (1.2), we use $s = |x^3| \ge 0$, increasing away from the plane on each side; the optical formulas apply separately to the two outgoing beams. For a single beam passing through a waist, one instead uses a signed axial coordinate, negative before the waist and positive after it. Its Gouy contribution is $-\psi(s)$ along that signed coordinate. In our source problem the contribution on both sides is $-\psi(|x^3|)$. These describe different configurations when comparing phases across the waist.

Figure 1 shows a single beam passing through its waist, together with its time-averaged profile.

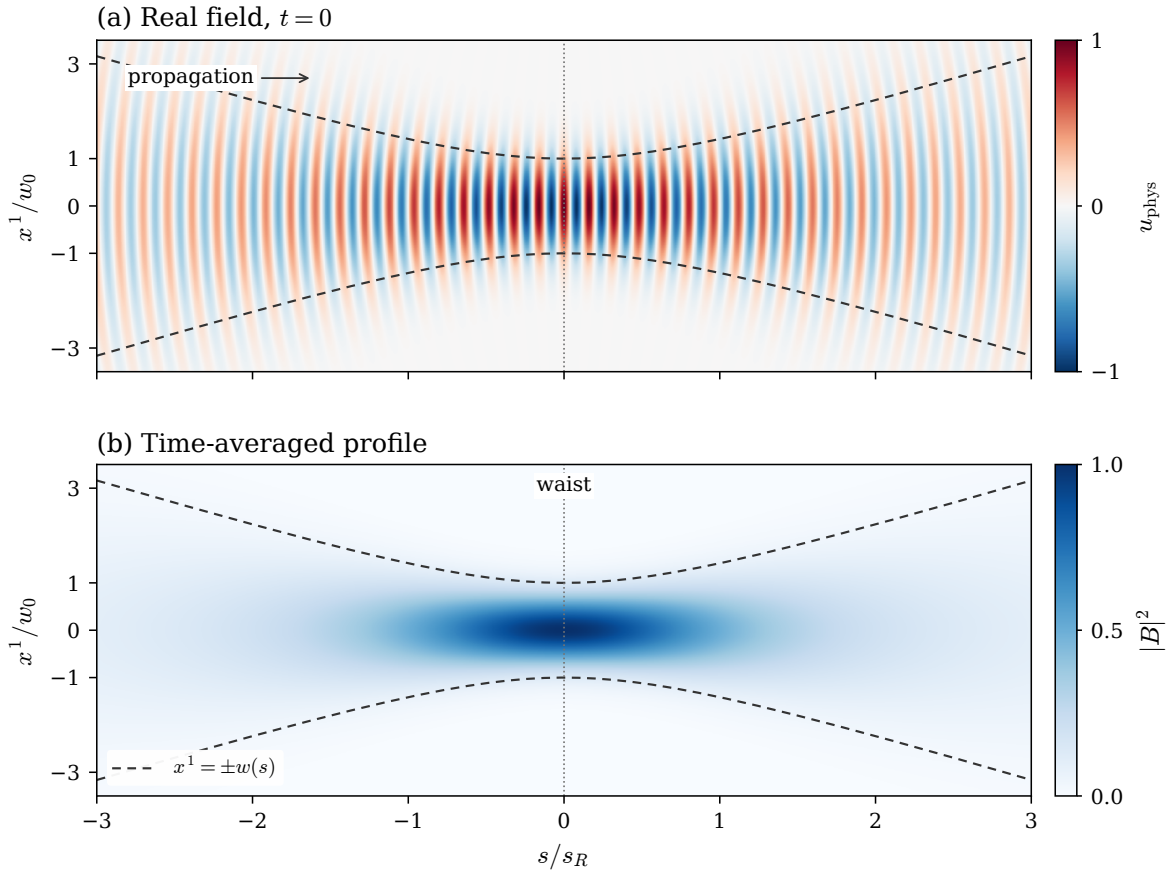


Figure 1. Both panels show the longitudinal cross-section $x^2 = 0$ of a three-dimensional Gaussian beam, with signed axial coordinate s and $\lambda s_R = 40$. The plots evaluate the explicit Gaussian formula. Its envelope $B(x', s)$ solves the free Schrödinger (paraxial) equation exactly in both transverse variables; the field $e^{\mathrm{i}\lambda s}B$ is a leading Gaussian-beam approximation for the full Helmholtz equation. No numerical solution of the full PDE is used. Top: the real field at $t = 0$. Bottom: $2\langle u^2_{\text{phys}}\rangle_t = |B|^2$, normalized to one at the waist center; this is the leading intensity profile. The dashed curves are $x^1 = \pm w(s)$, where the squared amplitude is e^{-2} times its on-axis value. Propagation is towards increasing s; the axes use different length scales.

The normalization in Figure 1 is fixed at the waist center, where $U(0, 0) = 1$. The upper panel plots the real field itself at $t = 0$. The mean squared field at the waist center is $1/2$, so dividing by this value gives the lower panel, $2\langle u^2_{\text{phys}}\rangle_t = |B|^2$. For the full three-dimensional beam,

$$|B(x', s)|^2 = \frac{w_0^2}{w(s)^2} e^{-2|x'|^2/w(s)^2}, \qquad \int_{\mathbb{R}^2} |B(x', s)|^2 \, \mathrm{d}x' = \frac{\pi w_0^2}{2}.$$

With the scalar energy normalization above, the leading power through a plane $s = \text{const}$ is therefore

$$P_{\text{lead}}(s) := \frac{\lambda^2}{2} \int_{\mathbb{R}^2} |B(x', s)|^2 \, \mathrm{d}x' = \frac{\pi \lambda^2 w_0^2}{4},$$

independent of s. The decrease of the on-axis squared amplitude is exactly balanced by the increase of the beam area. This conservation of the leading beam power is exact within the paraxial model. The integral is over both transverse variables $x' = (x^1, x^2)$, not just the slice $x^2 = 0$ shown in the figure.

We can compute the integrated correction as well. Evaluate the full scalar flux on the paraxial field $U = e^{\mathrm{i}\lambda s}B$. Since $\partial_s B = \mathrm{i}\Delta_{x'}B/(2\lambda)$, integration by parts gives

$$\begin{aligned} P_{\mathrm{GB}}(s) &:= \int_{\mathbb{R}^2} \langle J_s \rangle_t \, \mathrm{d}x' \\ &= P_{\mathrm{lead}} + \frac{\lambda}{2} \Im \int_{\mathbb{R}^2} \overline{B}\, \partial_s B \, \mathrm{d}x' \\ &= P_{\mathrm{lead}} - \frac{1}{4} \int_{\mathbb{R}^2} |\nabla_{x'} B|^2 \, \mathrm{d}x' = P_{\mathrm{lead}} - \frac{\pi}{4}. \end{aligned}$$

Here both $\int |B|^2 \, \mathrm{d}x'$ and $\int |\nabla_{x'} B|^2 \, \mathrm{d}x'$ are preserved by the free Schrödinger evolution. For the second integral we evaluate at the waist, where $B(x', 0) = e^{-|x'|^2/w_0^2}$:

$$\int_{\mathbb{R}^2} |\nabla_{x'} B(x', s)|^2 \, \mathrm{d}x' = \frac{4}{w_0^4} \int_{\mathbb{R}^2} |x'|^2 e^{-2|x'|^2/w_0^2} \, \mathrm{d}x' = \pi.$$

Thus P_{GB} is also independent of s. Its relative correction to P_{lead} is $-1/(\lambda^2 w_0^2)$, which is $-1/(2\lambda)$ for our choice $w_0^2 = 2/\lambda$. This is an exact integrated identity for the flux evaluated on the paraxial Gaussian. The field still has a Helmholtz residual, so its local wave-energy balance contains the residual term written above. To determine the next-order power of the exact Helmholtz solution, we must also include the higher order corrections to the beam itself.

The phase and transport equations. We can obtain this beam directly, without the Green's function. Following Ralston [12, Section I, pp. 1–4], look for a complex phase near the positive axis of the form

$$\phi_2(x', s) = s + \frac{1}{2} x' \cdot M(s) x', \qquad M(s)^T = M(s), \quad \Im M(s) > 0.$$

For a complex phase the eikonal equation is $\nabla\phi \cdot \nabla\phi = 1$, with no complex conjugation in the dot product. Direct substitution gives

$$\nabla\phi_2 \cdot \nabla\phi_2 - 1 = x' \cdot (M' + M^2) x' + \frac{1}{4} \big(x' \cdot M' x'\big)^2.$$

To cancel the quadratic term we solve the matrix Riccati equation

$$M' + M^2 = 0, \qquad M(0) = \mathrm{i} I_2.$$

It has the solution

$$M(s) = \frac{\mathrm{i}}{1 + \mathrm{i}s} I_2, \qquad \Im M(s) = \frac{1}{1 + s^2} I_2.$$

In particular, the imaginary part stays positive for every finite s. More generally, for symmetric M_0 with $\Im M_0 > 0$, we have

$$M(s) = M_0 (I + s M_0)^{-1}, \qquad \Im M(s) = (I + s M_0)^{-*} (\Im M_0) (I + s M_0)^{-1},$$

where $B^{-*} = (B^{-1})^*$. To see that $I + sM_0$ is invertible for real $s \neq 0$, suppose $(I + sM_0)v = 0$. Then $v^* M_0 v = -|v|^2/s$ is real, contradicting $v^*(\Im M_0)v > 0$ unless $v = 0$. This is the elementary Euclidean form of Ralston's argument for the nonsingularity of the complex phase.

For the amplitude, substitution gives

$$(-\Delta - \lambda^2)(e^{\mathrm{i}\lambda\phi}a) = e^{\mathrm{i}\lambda\phi}\left\{\lambda^2(\nabla\phi\cdot\nabla\phi - 1)a - \mathrm{i}\lambda\big(2\nabla\phi\cdot\nabla a + (\Delta\phi)a\big) - \Delta a\right\}.$$

Taking $a = A(s)$ and evaluating the transport equation on the axis, we get

$$2A' + (\operatorname{tr} M)A = 0, \qquad A(0) = 1, \quad \text{hence} \quad A(s) = \frac{1}{1+\mathrm{i}s}.$$

For this circular beam, the connection with the optical parameters is

$$M(s) = \left(\frac{1}{R(s)} + \frac{2\mathrm{i}}{\lambda w(s)^2}\right) I_2, \qquad A(s) = \frac{w_0}{w(s)} e^{-\mathrm{i}\psi(s)}.$$

Thus the real part of the phase Hessian gives the wavefront curvature, its imaginary part gives the Gaussian decay, and the phase of the transport amplitude gives the Gouy shift. In fact, the transport equation implies

$$\frac{\mathrm{d}}{\mathrm{d}s}\arg A(s) = -\frac{1}{2}\operatorname{tr}(\Im M(s)) = -\psi'(s).$$

With $A(0) = 1$ and the argument chosen continuously, this gives $\arg A(s) = -\psi(s) = -\arctan s$ in our normalization, in agreement with the outgoing convention.

This recovers u_{GB}. The value $A(0) = 1$ is the leading normalization furnished by our source; the stationary phase estimate itself was made away from that source. The quadratic phase solves the eikonal equation up to $O(|x'|^4)$, and this amplitude solves the transport equation on the axis. Thus this is the leading beam, not yet an approximation to arbitrary order. We explain the higher order corrections below.

The power of the prefactor depends on the number of transverse variables. With d transverse variables and the same normalization $M(0) = \mathrm{i}I_d$, the transport equation gives

$$A(s) = (1+\mathrm{i}s)^{-d/2}, \qquad \arg A(s) = -\frac{d}{2}\arctan s,$$

with the power and argument chosen continuously from $A(0) = 1$. Each transverse variable contributes $-\frac{1}{2}\arctan s$ to the Gouy phase; here $d = 2$. Ralston's explicit example [12, Section I, p. 5] has one transverse variable, which explains its exponent $-1/2$. His amplitude is $(1 + 2\mathrm{i}at)^{-1/2}$, with $a > 0$; taking $a = 1/2$ gives $(1 + \mathrm{i}t)^{-1/2}$ and the phase $-\frac{1}{2}\arctan t$. The wave packet below uses the conjugate carrier, so its conjugate prefactor has the opposite phase.

To match the source in (1.2) to higher order, we also have to match the jump across $x^3 = 0$. The field is continuous there and

$$\partial_{x^3} u|_{0+} - \partial_{x^3} u|_{0-} = 2\mathrm{i}\lambda e^{-\lambda|x'|^2/2}.$$

Use the same outgoing beam on both sides, with $s = |x^3|$, and prescribe $\phi(x', 0) = \mathrm{i}|x'|^2/2$. The outgoing root of the eikonal equation at the plane is then $\phi_s(x', 0) = \sqrt{1 + |x'|^2}$. The jump condition becomes, to the order under consideration,

$$\phi_s a + \frac{1}{\mathrm{i}\lambda}\partial_s a = 1 \quad \text{at } s = 0.$$

In particular, $a_0(x', 0) = (1 + |x'|^2)^{-1/2}$, whose value on the axis is 1, and the successive initial amplitudes satisfy $\phi_s a_j = \mathrm{i}\partial_s a_{j-1}$ there. These identities prescribe the Taylor data at $x' = 0$ for the higher order construction. They also explain why the leading expression alone need not reproduce the source exactly.

2. The metric case.

Take first $c = 1$, so $g = g_0$, and consider $(-\Delta_g - \lambda^2)U = 0$ with the same time factor $e^{-\mathrm{i}\lambda t}$. The beam axis is now a unit speed geodesic $\gamma(s)$. Use Fermi coordinates (s, y), obtained from a parallel orthonormal frame normal to γ, with $y \in \mathbf{R}^{n-1}$. The quadratic beam has the form

$$U_{\mathrm{GB}}(s, y) = A(s)e^{\mathrm{i}\lambda\phi_2(s,y)}, \qquad \phi_2(s, y) = s + \frac{1}{2}y \cdot M(s)y, \qquad M^T = M, \quad \Im M > 0.$$

The eikonal and transport equations along the axis give

$$M' + M^2 + \mathcal{R} = 0, \qquad 2A' + (\operatorname{tr} M)A = 0.$$

Here $\mathcal{R}(s)$ is the transverse curvature matrix, with the sign convention that the Jacobi equation for infinitesimal separation of geodesics is $Y'' + \mathcal{R}Y = 0$. To reduce the Riccati equation to this linear matrix equation, write $M = Y'Y^{-1}$. The order of the factors matters. Wherever Y is invertible, differentiation gives

$$M' = Y''Y^{-1} - Y'Y^{-1}Y'Y^{-1} = Y''Y^{-1} - M^2,$$

and hence

$$M' + M^2 + \mathcal{R} = (Y'' + \mathcal{R}Y)Y^{-1}.$$

This is an exact substitution, also when $\mathcal{R}$ varies with s. Conversely, given a Riccati solution M on an interval containing 0, solve $Y' = MY$, $Y(0) = I$. Let us justify Liouville's formula without assuming invertibility. Put $d(s) = \det Y(s)$ and let $\operatorname{adj} Y$ be the adjugate, the transpose of the cofactor matrix. Since $Y \operatorname{adj} Y = dI$, differentiation of the determinant and cyclicity of the trace give

$$d' = \operatorname{tr}((\operatorname{adj} Y)Y') = \operatorname{tr}(Y(\operatorname{adj} Y)M) = d \operatorname{tr} M.$$

These identities hold even when Y is singular. Solving this scalar equation with $d(0) = 1$ gives

$$\det Y(s) = \exp\left(\int_0^s \operatorname{tr} M(\sigma)\,\mathrm{d}\sigma\right) \neq 0,$$

so Y is invertible on that interval. Differentiating gives the Jacobi equation and its initial data:

$$Y'' = (M' + M^2)Y = -\mathcal{R}Y, \qquad Y'(0) = M(0).$$

See Ma, Sahoo, and Salo [9, Section 6, proof of Lemma 6.3] for the intrinsic form of this reduction and the regularity argument below. They use the normalization $M_0 = \mathrm{i}I$ in a parallel orthonormal frame; the same argument below allows any complex symmetric M_0 with $\Im M_0 > 0$.

Let us see why the complex Hessian stays regular. Given $M_0^T = M_0$, $\Im M_0 > 0$, solve the linear system

$$Y' = Z, \qquad Z' = -\mathcal{R}Y, \qquad Y(0) = I, \quad Z(0) = M_0.$$

Since $\mathcal{R}$ is real and symmetric, differentiation gives the conserved identities

$$Y^*Z - Z^*Y = 2\mathrm{i}\,\Im M_0, \qquad Y^TZ - Z^TY = 0.$$

If $Y(s)v = 0$, the first identity gives $v^*(\Im M_0)v = 0$, hence $v = 0$. Thus Y is invertible, and we can set

$$M = ZY^{-1}, \qquad \Im M = Y^{-*}(\Im M_0)Y^{-1} > 0.$$

The second conserved identity gives $M^T = M$. By continuity and compactness, $\Im M$ has a uniform positive lower bound on each fixed compact ray segment. On the same segment M is smooth and bounded; the transport ODE then keeps A regular as well. This explains regularity of the beam

coefficients through a real caustic. Coordinate patches may still be needed, and the full propagator kernel can remain singular there.

The amplitude can also be written directly in terms of Y. Since

$$\operatorname{tr} M = \operatorname{tr}(Y'Y^{-1}) = \frac{(\det Y)'}{\det Y},$$

the transport equation gives

$$A(s) = A(0)(\det Y(s))^{-1/2}.$$

The square root is continued continuously along the ray from $(\det Y(0))^{1/2} = 1$, since $Y(0) = I$. We keep this continuous branch through real caustics as well.

Focusing. The beam can still narrow and its amplitude can increase. Put $Q(s) = \Im M(s)$. The squared amplitude of the quadratic beam is

$$|U_{\mathrm{GB}}(s, y)|^2 = |A(s)|^2 e^{-\lambda y \cdot Q(s) y}.$$

If $q_j(s)$ are the eigenvalues of $Q(s)$, the transverse widths in the principal directions are of order $(\lambda q_j(s))^{-1/2}$. Thus increasing q_j means focusing in that direction. Conservation of leading beam power, verified below, makes the peak squared amplitude $|A|^2$ increase when the product of the transverse widths decreases. On a fixed compact ray segment we have $\delta I \le Q(s) \le CI$ for some $\delta, C > 0$. The widths therefore remain comparable to $\lambda^{-1/2}$, although their proportionality factors can vary appreciably. They do not collapse to zero.

A conjugate point need not make an individual Gaussian beam narrower. On the unit sphere, $\mathcal{R} = I$. Taking $M_0 = \mathrm{i}I$ and $A(0) = 1$, the equations above give

$$Y(s) = e^{\mathrm{i}s} I, \qquad M(s) = \mathrm{i}I, \qquad A(s) = e^{-\mathrm{i}(n-1)s/2}.$$

Thus $Q = I$ and $|A| = 1$: the leading transverse width and amplitude stay constant, even through the antipodal conjugate point at $s = \pi$. The real Jacobi matrix for geodesics starting at the same point is instead $\mathcal{J}(s) = \sin s\, I$, with $\mathcal{J}(0) = 0$, $\mathcal{J}'(0) = I$, and $\mathcal{J}(\pi) = 0$. Conjugacy describes focusing of that family of geodesics. A Gaussian beam has both spatial width and angular spread; its focusing also depends on the initial complex Hessian.

Let us compare S^2 and S^3. Here S^2 is the ordinary spherical surface, while S^3 has intrinsic dimension three. A ray is one-dimensional, so there is one transverse direction on S^2 and two on S^3. We can picture the beam as a narrow ribbon around a great circle on S^2, and as a tube with a two-dimensional cross-section on S^3. For the matched beam above, each transverse direction contributes $e^{-\mathrm{i}s/2}$ to the amplitude. Thus

$$\begin{aligned} S^2: &\quad A(s) = e^{-\mathrm{i}s/2}, \qquad A(\pi) = -\mathrm{i}, \\ S^3: &\quad A(s) = e^{-\mathrm{i}s}, \qquad A(\pi) = -1. \end{aligned}$$

These are phase shifts of $-\pi/2$ and $-\pi$, respectively, relative to the longitudinal carrier $e^{\mathrm{i}\lambda s}$. They accumulate smoothly from $s = 0$ to $s = \pi$; the amplitude does not jump at the conjugate point. Geometrically, one independent angular variation of rays from a point focuses at the antipode on S^2, and two on S^3. Each contributes the factor $-\mathrm{i}$. The focusing mechanism and the antipodal distance π are the same for both unit spheres; the matched Gaussian beam has constant width in both cases.

For this beam on S^3, $\det Y(s) = e^{2\mathrm{i}s}$ makes one full counterclockwise turn around zero as s runs from 0 to π. Its square root continued from 1 is $e^{\mathrm{i}s}$, which reaches -1 at the antipode. Thus

$A(\pi) = -1$ even though $\det Y(\pi) = 1$. Evaluating the principal square root anew at each point would lose this sign.

Higher transverse Taylor coefficients of the phase and amplitudes are obtained successively from ODEs along γ; the eikonal and transport equations are imposed to arbitrarily high order there.

After removing $e^{\mathrm{i}\lambda s}$ and retaining the leading terms on the beam scale $|y| = O(\lambda^{-1/2})$, the transverse model is still a Schrödinger equation, now with a quadratic potential:

$$\mathrm{i}\partial_s B = -\frac{1}{2\lambda}\Delta_y B + \frac{\lambda}{2}\, y \cdot \mathcal{R}(s) y\, B.$$

Substituting $B = A(s)e^{\mathrm{i}\lambda y\cdot M(s)y/2}$ gives precisely the two ODEs above. Thus curvature changes the spreading and focusing; the transverse Gaussian need not be circular.

Initial width and curvature. Keep the metric fixed as $\lambda \to \infty$. The beam shrinks to the central ray, but its profile on the transverse scale $\lambda^{-1/2}$ still evolves. Let us see this with one transverse variable and constant curvature $K = \kappa^2 > 0$. Take $M(0) = \mathrm{i}a$, $a > 0$, and $A(0) = 1$, with a and κ independent of λ. Setting $z = \sqrt{\lambda}\, y$ and $C(s, z) = B(s, z/\sqrt{\lambda})$ in the leading transverse equation gives

$$\mathrm{i}\partial_s C = -\frac{1}{2}\partial_z^2 C + \frac{\kappa^2 z^2}{2} C, \qquad C(0, z) = e^{-az^2/2}.$$

There is no λ left in this equation. It is the Schrödinger equation for the harmonic oscillator, $\mathrm{i}\partial_s C = H_\kappa C$, where

$$H_\kappa = -\frac{1}{2}\partial_z^2 + \frac{\kappa^2 z^2}{2}.$$

The corresponding classical Hamiltonian is $h(z, \zeta) = \frac{1}{2}\zeta^2 + \frac{1}{2}\kappa^2 z^2$. Hamilton's equations give $z' = \zeta$, $\zeta' = -\kappa^2 z$, hence $z'' + \kappa^2 z = 0$. This is the usual harmonic-oscillator ODE, also the transverse Jacobi equation here.

The special Gaussian with $a = \kappa$ is an eigenfunction of H_κ. Direct differentiation gives

$$H_\kappa e^{-\kappa z^2/2} = \frac{\kappa}{2} e^{-\kappa z^2/2}, \qquad C(s, z) = e^{-\mathrm{i}\kappa s/2} e^{-\kappa z^2/2}.$$

Its evolution changes only the phase, so its absolute value and width stay constant. For $a \neq \kappa$, the initial Gaussian is not this eigenfunction and its width varies, as we compute below. Diffraction and geometric focusing both remain in the leading rescaled dynamics. Spreading and focusing refer to changes of this profile, even though the physical width tends to zero everywhere along a fixed ray segment.

For general $a > 0$, let us look for a Gaussian solution of the Schrödinger equation in the form

$$C(s, z) = A(s) e^{\mathrm{i}M(s)z^2/2}, \qquad A(0) = 1, \quad M(0) = \mathrm{i}a.$$

Substitution gives

$$\mathrm{i}\frac{A'}{A} - \frac{1}{2}M' z^2 = -\frac{\mathrm{i}M}{2} + \frac{M^2 + \kappa^2}{2} z^2.$$

Equating the constant and quadratic terms, we obtain

$$2A' + MA = 0, \qquad M' + M^2 + \kappa^2 = 0.$$

To reduce the Riccati equation to a linear second-order equation, write $M = Y'/Y$, with $Y(0) = 1$. This is an exact substitution wherever $Y \neq 0$. Since $M' + M^2 = Y''/Y$, it becomes

$$Y'' + \kappa^2 Y = 0, \qquad Y(0) = 1, \quad Y'(0) = \mathrm{i}a.$$

Thus Y solves the classical oscillator equation with complex initial data. We get

$$Y(s) = \cos(\kappa s) + \frac{\mathrm{i}a}{\kappa}\sin(\kappa s), \qquad M = \frac{Y'}{Y}, \qquad A = Y^{-1/2},$$

where the amplitude equation gives the last identity, with the square root continued from $A(0) = 1$. More explicitly,

$$A(s) = \exp\left(-\frac{1}{2}\int_0^s \frac{Y'(\sigma)}{Y(\sigma)}\,\mathrm{d}\sigma\right) = |Y(s)|^{-1/2}e^{-\mathrm{i}\theta(s)/2},$$

where θ is the continuous argument of Y, with $\theta(0) = 0$. Since $a > 0$, Y never vanishes. As s goes from 0 to π/κ, it travels from 1 to -1 through the upper half-plane. Thus at the antipode,

$$\theta(\pi/\kappa) = \pi, \qquad A(\pi/\kappa) = -\mathrm{i}.$$

We continue the same branch past this point; choosing the principal square root separately on either side would introduce a sign jump. The factor $-\mathrm{i}$ is the transverse amplitude phase, in addition to the carrier $e^{\mathrm{i}\lambda s}$. Consequently,

$$C(s,z) = Y(s)^{-1/2}\exp\left(\frac{\mathrm{i}}{2}\frac{Y'(s)}{Y(s)}z^2\right).$$

This solves the quadratic Schrödinger model exactly. For the original beam, it describes the leading transverse behavior. Thus Y determines the Gaussian width, wavefront curvature, and amplitude. In particular,

$$\Im M(s) = \frac{a}{\cos^2(\kappa s) + (a^2/\kappa^2)\sin^2(\kappa s)}.$$

Define $w_\lambda(s)$ by $|U_{\mathrm{GB}}(s, w_\lambda(s))|^2 = e^{-2}|U_{\mathrm{GB}}(s,0)|^2$, as in the optical convention above. Then

$$w_\lambda(s)^2 = \frac{2}{\lambda}\left(\frac{\cos^2(\kappa s)}{a} + \frac{a}{\kappa^2}\sin^2(\kappa s)\right).$$

Thus $\lambda^{-1/2}$ sets the small spatial scale, while the initial Hessian and the curvature determine its varying coefficient. If $a > \kappa$, the beam starts narrower than the stationary Gaussian and spreads up to $s = \pi/(2\kappa)$. If $a < \kappa$, it starts broader and narrows up to that point. For $a = \kappa$, $M(s) = \mathrm{i}\kappa$ and the width is constant. The Fourier transform of the initial Gaussian has transverse angular spread of order $\sqrt{a/\lambda}$, while its spatial width is of order $(\lambda a)^{-1/2}$. A narrower initial beam therefore has greater angular spread.

The conjugate distance is a separate geometric quantity. The real Jacobi field with $j(0) = 0$, $j'(0) = 1$ is

$$j(s) = \frac{\sin(\kappa s)}{\kappa},$$

so the first conjugate point occurs at $s = \pi/\kappa$, independently of a and λ. More generally, the linearized transverse motion is

$$y(s) = \cos(\kappa s)y_0 + \frac{\sin(\kappa s)}{\kappa}\eta_0,$$

where y_0 and η_0 are the initial transverse displacement and angular variation. At the conjugate point the dependence on η_0 disappears, but $y(\pi/\kappa) = -y_0$. A Gaussian beam has initial spatial

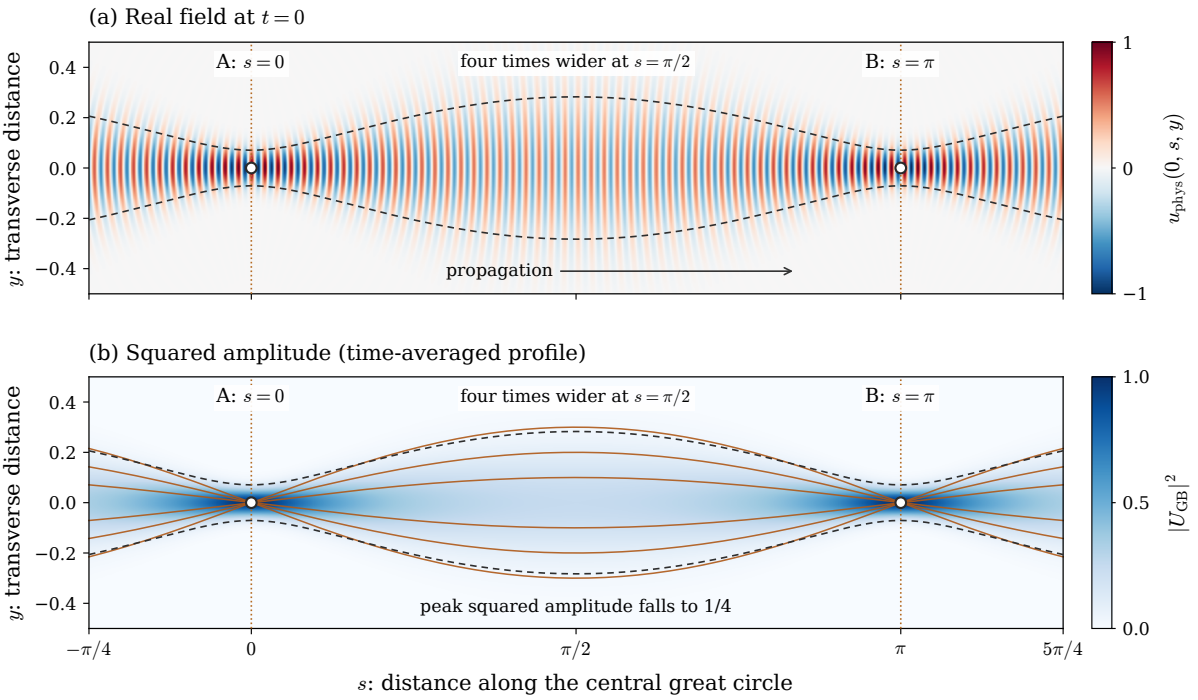


Figure 2. Spreading and refocusing of a leading Gaussian beam on the unit two-sphere, with $M(0) = 4\mathrm{i}$, $A(0) = 1$, and $\lambda = 100$. The plots evaluate the explicit leading quadratic beam. After removing $e^{\mathrm{i}\lambda s}$ and setting $z = \sqrt{\lambda}\, y$, its envelope $C(s, z)$ solves the harmonic-oscillator Schrödinger equation exactly. The full spherical Helmholtz equation is satisfied only approximately; no numerical solution of it is used. The flat strip has one transverse coordinate y and metric $\mathrm{d}y^2 + \cos^2 y\, \mathrm{d}s^2$. Top: the real field at $t = 0$, with time factor $e^{-\mathrm{i}\lambda t}$. Bottom: $|U_{\mathrm{GB}}|^2 = 2\langle u_{\mathrm{phys}}^2 \rangle_t$, with fixed normalization $|U_{\mathrm{GB}}(0, 0)|^2 = 1$. Dashed curves: $y = \pm w_\lambda(s)$. Orange curves: exact geodesics from A to its antipode B, shown for comparison.

width as well as angular spread. It therefore returns to its initial width there, without collapsing to a point. For $a > \kappa$ this is refocusing after spreading; for $a = \kappa$ the width never changed.

Figure 2 shows the case $\kappa = 1$, $a = 4$. The width at $s = \pi/2$ is four times its value at $s = 0$, and the peak squared amplitude falls to $1/4$. At the antipode $s = \pi$ both return to their initial values. The overlaid geodesics start at the single point A and focus at B; they illustrate conjugacy, while the beam retains a nonzero width. In these coordinates the geodesic leaving A at angle α to the central great circle is $y(s) = \arctan(\tan\alpha \sin s)$.

The leading power has the same geometric interpretation. Set $d = n - 1$. The ODEs imply $Q' = -(\Re M)Q - Q(\Re M)$ and hence $|A|^2/\sqrt{\det Q}$ is constant. For $v = \Re(e^{-\mathrm{i}\lambda t}U)$ the scalar flux is $J = -v_t \nabla_g v$. Integrating its leading normal component across a local section normal to γ, with the Riemannian surface measure, gives

$$P_{\mathrm{lead}}(s) = \frac{\lambda^2}{2} \left(\frac{\pi}{\lambda}\right)^{d/2} \frac{|A(s)|^2}{\sqrt{\det Q(s)}}.$$

The surface measure has leading term $\mathrm{d}y$ in these coordinates, so the formula follows from the transverse Gaussian integral. This recovers the constant leading power above when $d = 2$ and

$\mathcal{R} = 0$. Dahl [2, Sections 1–3] gives a geometric treatment of the amplitude and the corresponding leading energy conservation for wave packets. The Euclidean width law and the correction $-\pi/4$ are not universal: curvature and higher order beam terms enter the corrections to the flux.

For the original equation (II.1.1) with variable c, the rays are geodesics of $g = c^{-2}g_0$. One must still use the transport equation of the actual operator $-c^2\Delta_{g_0}$, which generally differs from $-\Delta_g$. The physical flux is $-v_t\nabla_{g_0}v$, integrated with the surface measure of g_0.

For the high-frequency Helmholtz application of Cekić, Krupchyk, Sahoo, and Uhlmann, see Chapter II, Section 11.

3. The wave equation

The monochromatic beams above are localized transversely to their axes. Let us now construct a moving packet which is also localized in the direction of travel, following the example and construction in Ralston [12, Section I, pp. 1–5]. We start again with the Euclidean equation (1.1). Work in $\mathbf{R}^n$, $n \geq 2$, and take $x_0 = 0$, $\omega = e_n$. Write $x = (x', x^n)$ and $q = x^n - t$, so the ray is $x' = 0$, $q = 0$.

To agree with the sign in (0.1), we start with phase $t - x^n$. This is the central phase of the $U_-(t)$ branch introduced in Chapter II, Section 10. Its initial spatial covector is $-\lambda e_n$. The canonical relation in Chapter III, Section 2 transports it by Φ_{-t}: in the Euclidean case, $(0, -\lambda e_n)$ goes to $(te_n, -\lambda e_n)$. Thus the packet center moves towards $+e_n$. We add imaginary quadratic terms:

$$\phi_2(t,x) = -q + \frac{1}{2}x' \cdot M(t)x' + \frac{\mathrm{i}}{2}q^2, \qquad M(0) = \mathrm{i}I_{n-1}.$$

The last term localizes the packet in the direction of travel as well. Put

$$E(\phi) = \nabla_x\phi \cdot \nabla_x\phi - \phi_t^2.$$

A direct calculation gives

$$E(\phi_2) = x' \cdot (M^2 - \dot{M})x' + \mathrm{i}q\, x' \cdot \dot{M}x' - \frac{1}{4}\big(x' \cdot \dot{M}x'\big)^2.$$

We therefore choose $\dot{M} = M^2$. The sign differs from the stationary calculation because the central phase here is $t - x^n$. We get

$$M(t) = \frac{\mathrm{i}}{1 - \mathrm{i}t}I_{n-1}, \qquad \Im\phi_2(t,x) = \frac{|x'|^2}{2(1+t^2)} + \frac{(x^n - t)^2}{2}.$$

Let $\Box = \partial_t^2 - \Delta$ as before, and set

$$T_\phi = 2(\phi_t\partial_t - \nabla_x\phi \cdot \nabla_x) + \Box\phi.$$

Formula (II.7.3) applies to complex phases as well, with $|\nabla_x\phi|^2$ there interpreted as $\nabla_x\phi \cdot \nabla_x\phi = \sum_j(\partial_j\phi)^2$, without complex conjugation. On the ray, $\phi_t = 1$ and $\Box\phi_2 = -\operatorname{tr} M$. Thus the leading amplitude $A(t)$ satisfies

$$2\dot{A} - (\operatorname{tr} M)A = 0, \qquad A(0) = 1, \quad \text{so} \quad A(t) = (1 - \mathrm{i}t)^{-(n-1)/2}.$$

The power is chosen continuously from $A(0) = 1$. The resulting leading packet is

$$u_{\mathrm{GB}}(t,x) = (1 - \mathrm{i}t)^{-(n-1)/2} \exp\left(\mathrm{i}\lambda(t - x^n) - \frac{\lambda|x'|^2}{2(1 - \mathrm{i}t)} - \frac{\lambda(x^n - t)^2}{2}\right).$$

At $t = 0$ it is the Gaussian version of (0.1), namely $e^{-\mathrm{i}\lambda x^n - \lambda|x|^2/2}$. Its absolute value is

$$|u_{\mathrm{GB}}(t,x)| = (1+t^2)^{-(n-1)/4} \exp\left(-\frac{\lambda|x'|^2}{2(1+t^2)} - \frac{\lambda(x^n-t)^2}{2}\right).$$

We see the center moving along $x = te_n$, with width of order $\lambda^{-1/2}$ in all spatial directions for bounded t. Translations and rotations give the packet along $x = x_0 + t\omega$. Complex conjugation changes the central phase to $x^n - t$ and gives the positive-frequency convention used for the outgoing Helmholtz wave. Time reversal gives the packet traveling in the opposite direction.

As in Chapter II, Section 7, the initial velocity is part of the construction. For the packet just displayed it is

$$(u_{\mathrm{GB}})_t(0,x) = \left[\mathrm{i}\lambda\left(1 - \frac{|x'|^2}{2}\right) + \lambda x^n + \frac{\mathrm{i}(n-1)}{2}\right] u_{\mathrm{GB}}(0,x).$$

Keeping only $\mathrm{i}\lambda u_{\mathrm{GB}}(0,x)$ would discard terms which matter when estimating the error.

An exact pure U_- half-wave must have both Cauchy data related by

$$u(t) = U_-(t)f_\lambda = e^{\mathrm{i}tB} f_\lambda, \qquad u_t(0) = \mathrm{i}Bf_\lambda.$$

Here $B = \sqrt{-\Delta}$; for the variable coefficient equation use $B = L^{1/2}$ as in Chapter II, Section 10. The Gaussian packet above is a leading approximation on this branch. Identifying its central phase does not assert the exact identity $u_{\mathrm{GB}}(t) = U_-(t)u_{\mathrm{GB}}(0)$, or that its differentiated initial velocity equals $\mathrm{i}Bu_{\mathrm{GB}}(0)$ exactly.

Higher order corrections. The formulas above describe the leading packets. For the wave packet put $d = (|x'|^2 + q^2)^{1/2}$. We have $E(\phi_2) = O(d^3)$ and $T_{\phi_2}A = O(d)$ near the ray. In particular, the quadratic phase and the amplitude $A(t)$ alone do not make the residual small to every order: on a fixed time interval their residual is bounded pointwise by $C\lambda^{1/2}$ for $\lambda \geq 1$.

The way to improve this is to continue the Taylor expansions along the ray, as in Ralston [12, Section I, pp. 3–4]. After the quadratic terms have been found, differentiating the eikonal equation gives linear ODEs along the ray for the higher derivatives of ϕ; their coefficients and right hand sides involve the derivatives already determined. We solve these equations successively. We then take

$$a(t,x,\lambda) \sim \sum_{j=0}^{\infty} \lambda^{-j} a_j(t,x), \qquad T_\phi a_0 = 0, \quad T_\phi a_j = \mathrm{i}\Box a_{j-1}, \quad j \geq 1,$$

where the transport equations, too, are imposed to the required Taylor order on the ray. Differentiating them gives linear ODEs for the derivatives of the amplitudes. The function $A(t)$ above specifies only the value of a_0 on the ray; its transverse derivatives must also be constructed. The stationary Helmholtz beam is improved in the same way using its eikonal and transport equations.

We are prescribing derivatives at the ray. To any finite order they can be realized by Taylor polynomials in coordinates transverse to it; prescribing all orders and taking a smooth realization makes the eikonal and transport errors vanish to infinite order there. This does not assert that the complex eikonal equation is solved exactly in a whole neighborhood. On a fixed ray segment the positive imaginary quadratic part gives $\Im\phi \geq cd^2$ in a sufficiently small tube. The elementary estimate

$$d^m e^{-c\lambda d^2} \leq C_m \lambda^{-m/2}$$

then turns vanishing to high order on the ray into a small error in λ. Multiply by a smooth cutoff equal to one in a smaller tube. Its derivatives are supported away from the ray and give

exponentially small errors. By retaining sufficiently many phase derivatives and amplitude terms, for any prescribed K we obtain a packet u_{app} such that

$$\sup_{|t|\le T} \|\Box u_{\mathrm{app}}(t)\|_{L^2} = O(\lambda^{-K}).$$

Let u be the exact wave with both Cauchy data equal to those of u_{app}. The energy estimate used in Chapter II, Section 7 gives

$$\sup_{|t|\le T} \left(\|u(t) - u_{\mathrm{app}}(t)\|_{H^1} + \|u_t(t) - \partial_t u_{\mathrm{app}}(t)\|_{L^2}\right) = O(\lambda^{-K}).$$

Thus the beam construction gives actual waves concentrated near the chosen ray to arbitrary accuracy on each fixed finite time interval. All these waves are smooth for fixed λ; their concentration is an asymptotic statement as $\lambda \to \infty$.

The wave equation with a metric. For (II.1.1), replace the straight ray by a unit speed geodesic $\gamma(t)$ of $g = c^{-2} g_0$. In local coordinates moving with $\gamma(t)$, repeat the Taylor construction, using the wave eikonal equation (II.2.11) and the transport equations of $\partial_t^2 - c^2 \Delta_{g_0}$. The quadratic phase is determined by a matrix Riccati equation associated with the linearized Hamilton flow; positive definite initial imaginary part gives localization near $x = \gamma(t)$. Higher phase derivatives and amplitude coefficients again satisfy ODEs along the ray. The finite-time error argument uses the energy norm of Chapter II, Section 1 and gives actual waves close to these packets. For $c = 1$, Dahl [2, Sections 1–3] describes the geometric amplitude and leading energy conservation. The choice of phase branch is made from the Cauchy data, just as in the Euclidean example.

More generally, the construction extends to scalar differential operators of real principal type, with geodesics replaced by null bicharacteristics of the principal symbol. Oksanen, Salo, Stefanov, and Uhlmann [11, Section 3, Theorem 3.1] construct quasimodes (approximate solutions) concentrating along injective null bicharacteristic segments as integrals of Gaussian wave packets. This also handles cusps of the spatial projection, where the ordinary beam coordinates break down. The amplitude transport includes the lower order terms of the operator; a positive conserved wave energy as above does not follow from the real principal type condition alone.

Chapter V

The wave equation in a domain with boundary

1. The 1D case

Consider the constant coefficient equation (II.1.3) on an interval $x \in (a, b)$, where we allow either endpoint to be infinite. We will review the solution formulas for various problems. In this chapter g denotes the background spatial metric (g_0 in Chapter II). Thus for the operator $\partial_t^2 - c^2\Delta_g$ the travel-time metric is $c^{-2}g$.

1.1. A boundary source. Let $(a, b) = (0, \infty)$.

Dirichlet boundary condition. Assume we have a boundary source $u|_{x=0} = f(t)$ at $x = 0$, i.e., we are solving

$$\begin{cases} u_{tt} - c^2 u_{xx} = 0, & t \in \mathbf{R},\ x > 0, \\ u|_{x=0} = f(t). \end{cases} \tag{1.1}$$

The solution then is

$$u = F(t - x/c) + G(t + x/c) \tag{1.2}$$

for any two functions F, G so that $f = F + G$. One can interpret this in the following way: we can split the boundary source f in two parts anyway we want; then one of the parts, $F(t - x/c)$, moves to the right when time increases, hence we can call it outgoing. The other part, $G(t + x/c)$ moves to the left, and we will call it incoming. Since there is nothing to the left of the boundary, this may sound counter-intuitive. We may think of $G(t + x/c)$ as back-propagated from the boundary; or in other words, it is that wave propagating to the left which has boundary trace $G(t)$.

Suppose $f = 0$ for $t < 0$. To select the wave emitted by this source, prescribe zero Cauchy data at $t = 0$ and solve forward in time,

$$\begin{cases} u_{tt} - c^2 u_{xx} = 0, & t > 0,\ x > 0, \\ u|_{x=0} = f(t), \\ u|_{t=0} = 0, \\ u_t|_{t=0} = 0. \end{cases} \tag{1.3}$$

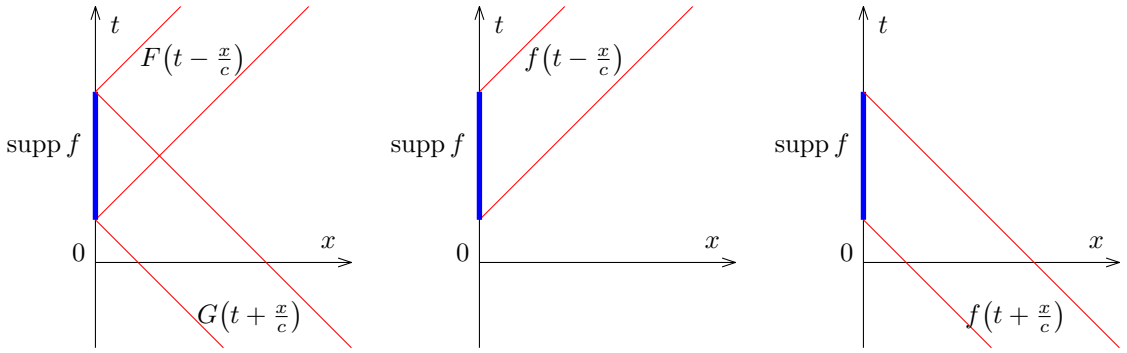


Figure 1. Propagation of a boundary source f on the half-line $x > 0$. Left: general solution, where $F + G = f$. Center: outgoing solution. Right: incoming solution.

What do the initial conditions tell us about the two profiles in (1.2)? We have $F(-x/c)+G(x/c) = 0$ and $F'(-x/c) + G'(x/c) = 0$. Differentiating the first equality in x gives $-F'(-x/c) + G'(x/c) = 0$ as well. Thus both profiles are constant on the arguments reached at $t = 0$. We can take those constants to be zero since we may add a constant to F and subtract it from G. For $t > 0$ we now have $G(t + x/c) = 0$, and the boundary condition gives $u_{\text{out}} = f(t - x/c)$. This solution extends by zero to $t < 0$.

The time origin plays no special role. If $f = 0$ for $t < a$, prescribe zero Cauchy data at $t = a$ instead. The same formula gives the retarded solution, which vanishes for $t < a$. If the source acted before $t = 0$, this wave can already be in the half-line at $t = 0$. Similarly, if $f = 0$ for $t > b$, requiring the wave to vanish for $t > b$ gives the incoming solution $u_{\text{in}} = f(t + x/c)$.

Neumann boundary condition. Assume we have a boundary source for u_x, i.e., $u_x|_{x=0} = f(t)$ at $x = 0$. In other words, we are solving

$$\begin{cases} u_{tt} - c^2 u_{xx} = 0, & t \in \mathbf{R},\ x > 0, \\ u_x|_{x=0} = f(t). \end{cases} \tag{1.4}$$

Differentiating (1.2) at $x = 0$ gives $-F'(t) + G'(t) = cf(t)$. We can still choose one profile freely. Once we select a single branch, its profile is determined up to an additive constant. For the outgoing wave $F' = -cf$. If $f = 0$ for $t \ll 0$, requiring the wave to vanish in the past gives $u_{\text{out}}(t, x) = -c\int_{-\infty}^{t-x/c} f(s)\,\mathrm{d}s$. For the incoming wave $G' = cf$. If $f = 0$ for $t \gg 0$, requiring the wave to vanish in the future gives $u_{\text{in}}(t, x) = -c\int_{t+x/c}^{\infty} f(s)\,\mathrm{d}s$.

The outgoing and the incoming DN maps. For a Dirichlet source f vanishing in the distant past, take the retarded solution above and define

$$\Lambda_{\text{out}} f(t) = -\frac{\partial u_{\text{out}}}{\partial x}\Big|_{x=0}, \tag{1.5}$$

where the minus sign comes from the outward unit normal to the half-line. Similarly, for f vanishing in the distant future, use the incoming solution to define Λ_{in}. Differentiating the two profiles gives

$$\Lambda_{\text{out}} = \frac{1}{c}\frac{\mathrm{d}}{\mathrm{d}t} = -\Lambda_{\text{in}}. \tag{1.6}$$

1.2. Reflections from the boundary. Let $(a, b) = (0, \infty)$.

Homogeneous Dirichlet boundary condition. We are solving

$$\begin{cases} u_{tt} - c^2 u_{xx} = 0, & t \in \mathbf{R},\ x > 0, \\ u|_{x=0} = 0. \end{cases} \tag{1.7}$$

which has infinitely many solutions. Solutions of such problems with prescribed Cauchy data at a certain time are given usually by taking odd/even extensions of the Cauchy data and reducing this to a problem on the whole line. In this case, we get directly

$$u = F(t + x/c) - F(t - x/c), \quad x > 0 \tag{1.8}$$

for an arbitrary F. One can interpret the results as follows: the first term is a wave incident on the boundary $x = 0$. When a part of it passes through $x = 0$ into $x < 0$, that part reflects back (x gets replaced by $-x$) and changes sign. This creates a reflection going away from $x = 0$. Another way to look at (1.8) is to think about $u_{\text{in}} = F(t + x/c)$, say with F supported in $(0, \infty)$, approaching $x = 0$. We want to satisfy the homogeneous Dirichlet boundary condition at $x = 0$ by sending an outgoing wave $u_{\text{out}} = G(t - x/c)$ so that $u = u_{\text{in}} + u_{\text{out}}$ solves (1.7). At $x = 0$ the incident wave has trace $F(t)$, so the reflected wave must have boundary data $u_{\text{out}}(t, 0) = -F(t)$ to make the total trace zero. Taking $f(t) = -F(t)$ in the Dirichlet source problem (1.1) gives the outgoing solution $u_{\text{out}}(t, x) = -F(t - x/c)$.

Homogeneous Neumann boundary condition. We are solving

$$\begin{cases} u_{tt} - c^2 u_{xx} = 0, & t \in \mathbf{R},\ x > 0, \\ u_x|_{x=0} = 0. \end{cases} \tag{1.9}$$

which has infinitely many solutions again. An elementary calculation shows that the general solution is given by

$$u = F(t + x/c) + F(t - x/c), \quad x > 0, \tag{1.10}$$

compare with the Dirichlet reflection in (1.8). In other words, to satisfy the homogeneous Neumann condition, we send the reflected wave back with the same sign. If we think of the reflected wave as a wave $u_{\text{out}} = G(t - x/c)$ related to a unknown G, then the Dirichlet data for those waves are F and G, respectively, and since u of the form (1.2) must satisfy $u_x = 0$ at $x = 0$, we get

$$\Lambda_{\text{in}} F + \Lambda_{\text{out}} G = 0,$$

which simply means that $F' = G'$. Then $F = G + C$ with some constant C and setting $\tilde{F} = F - C/2$, $\tilde{G} = G + C/2$, we get (1.10) with $\tilde{F}$ instead of F but still arbitrary.

2. The wave equation with a boundary source. Incoming and outgoing solutions

Let (f_1, f_2) be supported in a compact set $K \subset \Omega$, where Ω is a bounded domain with smooth boundary. We want to solve the wave equation (II.1.1) with homogeneous Dirichlet or Neumann boundary conditions. The boundary condition determines the self-adjoint realization of $-c^2\Delta_g$ in $L^2(\Omega, c^{-2}\mathrm{d}\,\mathrm{Vol}_g)$. The corresponding first order wave generator $\mathbf{P}$ in (II.1.11) is skew-adjoint on the energy space. In the Neumann case the energy determines the displacement up to an additive constant in space.

By finite speed of propagation (see the cone argument in Evans), for $|t| \ll 1$ the solution is still supported away from $\partial\Omega$, so the interior parametrix applies. If smooth data reach the boundary, we must also check compatibility at the corner. For homogeneous Dirichlet data, for example,

$f_1|_{\partial\Omega} = f_2|_{\partial\Omega} = 0$, and the higher order conditions follow by differentiating the boundary condition in time and using the equation. Even a smooth f_1 equal to one near $\partial\Omega$ violates the first condition.

Consider the mixed problem for the wave equation in Ω

$$\begin{cases} (\partial_t^2 - c^2(x)\Delta_g)u = 0 & \text{in } (0,T)\times\Omega, \\ u|_{(0,T)\times\partial\Omega} = h, \\ u|_{t=0} = 0, \\ \partial_t u|_{t=0} = 0. \end{cases} \tag{2.1}$$

With general smooth Cauchy data the first two compatibility conditions are $h(0,\cdot) = f_1|_{\partial\Omega}$ and $\partial_t h(0,\cdot) = f_2|_{\partial\Omega}$. For (2.1) the Cauchy data vanish, so smooth compatibility to all orders means $\partial_t^k h(0,\cdot) = 0$ for every $k \geq 0$. We take h supported away from $t = 0$. If we also have Cauchy data supported in the interior as above, we can add their solution with homogeneous boundary condition.

How much of the solution do we want to construct? Take h supported near (t_0, x_0) on the boundary and follow the wave as it first enters Ω. On an interval of length L with constant speed c, a pulse returns to its starting end after time $2L/c$ and contributes again to the normal trace. Here we work near the source, before any reflected ray returns to the patch. It is this local outgoing wave that we use later in the transmission problem.

Choose boundary normal coordinates (x', x^n) near x_0, with $x^n > 0$ in Ω. We take x^n to be distance to $\partial\Omega$ in the background metric g, so that for g Euclidean we can still keep c explicit. We look for a solution of the form

$$u(t,x) = (2\pi)^{-n} \iint e^{\mathrm{i}\phi(t,x,\tau,\xi')} a(t,x,\tau,\xi') \hat{h}(\tau,\xi')\, \mathrm{d}\tau\, \mathrm{d}\xi', \tag{2.2}$$

where $\hat{h}$ is the Fourier transform of $h(t,x')$ in both variables. We are prescribing data on a different hypersurface now. To recover h by Fourier inversion at $x^n = 0$, we take

$$\phi|_{x^n=0} = t\tau + x'\cdot\xi'. \tag{2.3}$$

As in (II.3.22), we can separate the time variable by writing $\phi = t\tau + \psi(x,\tau,\xi')$. Then $c^2|\nabla_x\psi|_g^2 = \tau^2$, with $\psi = x'\cdot\xi'$ on $x^n = 0$. We will return to the initial conditions for the amplitude later.

2.1. Constant coefficients, flat geometry. Let g be Euclidean and take $c = 1$ first. We try the phase

$$\phi = t\tau + x'\cdot\xi' + x^n\xi_n, \quad \xi_n^2 = \tau^2 - |\xi'|^2.$$

The transport equation allows us to take the amplitude equal to one. For each choice of root we get a solution of the form

$$u_\pm(t,x) = (2\pi)^{-n} \iint e^{\mathrm{i}\phi_\pm(t,x,\tau,\xi')} \hat{h}(\tau,\xi')\, \mathrm{d}\tau\, \mathrm{d}\xi'. \tag{2.4}$$

Which root gives the outgoing wave? Recall what we mean by outgoing here. As t increases, the wave leaves the boundary and enters $x^n > 0$. In the hyperbolic region $|\tau| > |\xi'|$ the normal velocity along a ray is $\mathrm{d}x^n/\mathrm{d}t = -\xi_n/\tau$. Thus ξ_n has the opposite sign to τ for the outgoing wave and the same sign for the incoming one.

In the elliptic region $|\tau| < |\xi'|$, both constructions use the root with positive imaginary part. Indeed, the factor $e^{\mathrm{i}x^n\xi_n}$ then decays as we move into the domain. Writing ϕ_+ for the outgoing

phase and ϕ_- for the incoming one, we get

$$\phi_\pm = t\tau + x'\cdot\xi' + x^n\xi_n^\pm, \quad \xi_n^\pm = \begin{cases} \mp\operatorname{sgn}(\tau)\sqrt{\tau^2-|\xi'|^2}, & |\tau|\ge|\xi'|, \\ \mathrm{i}\sqrt{|\xi'|^2-\tau^2}, & |\tau|<|\xi'|. \end{cases} \tag{2.5}$$

At glancing the roots meet at zero. The formulas are continuous there, although the square root loses smoothness. At normal incidence, restoring the constant speed c, we get $\xi_n^\pm = \mp\tau/c$. Applying $-\partial_{x^n}$ at the boundary therefore gives the multipliers $\pm\mathrm{i}\tau/c$, just as in (1.6).

Let us take the Fourier transform in (t, x') only, keeping $x^n \ge 0$ as a parameter. Formula (2.4) then reads

$$\widehat{u}_\pm(\tau,\xi',x^n) = e^{\mathrm{i}x^n\xi_n^\pm(\tau,\xi')}\hat{h}(\tau,\xi').$$

In a fixed spacelike direction the multiplier is $e^{-x^n\sqrt{|\xi'|^2-\tau^2}}$ for $c=1$. This is an evanescent wave. For any fixed $x^n > 0$ the multiplier decays exponentially as the frequency increases in a closed cone inside the elliptic region. At fixed temporal frequency $|\tau|$, the tangential frequency $|\xi'|$ can be arbitrarily large. Such fine spatial detail is accessible very close to the boundary, which is the point of using evanescent waves in near-field imaging.

2.2. Variable speed. Let us keep x^n equal to the distance to $\partial\Omega$ in the metric g, as above. In these coordinates $g^{nn}=1$ and $g^{n\alpha}=0$ for $\alpha = 1,\dots,n-1$. The eikonal equation becomes

$$\phi_t^2 - c^2\big(g^{\alpha\beta}\phi_{x^\alpha}\phi_{x^\beta} + \phi_{x^n}^2\big) = 0. \tag{2.6}$$

The boundary condition (2.3) fixes the tangential differential at (τ,ξ'). It remains to determine the normal derivative. At $x^n=0$ we have

$$(\partial_{x^n}\phi)^2 = \tau^2/c^2 - |\xi'|_g^2, \qquad |\xi'|_g^2 = g^{\alpha\beta}\xi_\alpha\xi_\beta,$$

with the coefficients evaluated at the boundary point. We get the same three cases as before.

(i) If $\tau^2/c^2 > |\xi'|_g^2$, there are two simple real roots. The normal ray velocity is $-c^2\partial_{x^n}\phi/\tau$. For the outgoing phase we take $\partial_{x^n}\phi = -\operatorname{sgn}(\tau)\sqrt{\tau^2/c^2-|\xi'|_g^2}$, and for the incoming phase we take the other root. The method of characteristics gives a local phase for each.

(ii) If $\tau^2/c^2 < |\xi'|_g^2$, the decaying choice is $\partial_{x^n}\phi = \mathrm{i}\sqrt{|\xi'|_g^2-\tau^2/c^2}$ at the boundary. For smooth coefficients we solve the eikonal equation and the successive transport equations to infinite order at $x^n=0$, as we explain below.

(iii) At $\tau^2/c^2 = |\xi'|_g^2$ the normal derivative vanishes and the ray is tangent to the boundary. Near such a covector the simple-root construction breaks down. This is the glancing case.

2.3. Evanescent waves. How far does an oscillation on the boundary penetrate into the medium? Take the flat half-space $x^n > 0$ with constant speed c, and prescribe a single boundary mode $h = e^{\mathrm{i}\lambda(t\tau+x'\cdot\xi')}$, with $\lambda > 0$. Write $u = h\,v(x^n)$ and plug this into $(\partial_t^2 - c^2\Delta)u = 0$. We get

$$v'' = \lambda^2\big(|\xi'|^2 - \tau^2/c^2\big)v, \qquad v(0)=1.$$

In the elliptic region $|\xi'| > |\tau|/c$, put $\kappa = \sqrt{|\xi'|^2-\tau^2/c^2}$. The general solution of the normal equation is a linear combination of $e^{\lambda\kappa x^n}$ and $e^{-\lambda\kappa x^n}$. Requiring boundedness as $x^n\to+\infty$ selects the decreasing exponential. Thus

$$u = e^{\mathrm{i}\lambda(t\tau+x'\cdot\xi')}e^{-\lambda\kappa x^n}. \tag{2.7}$$

The amplitude falls by a factor of e over the distance $1/(\lambda\kappa)$. For a fixed elliptic direction this distance shrinks like λ^{-1}. As we approach glancing, κ tends to zero and the decay becomes weaker. With the outward normal convention of (1.5), (2.7) gives $\partial_\nu u|_{x^n=0} = \lambda\kappa h$.

For a curved boundary or variable speed the same calculation gives the initial rate of decay. Take a real boundary phase φ and write $(\tau, \xi') = (\partial_t\varphi, \mathrm{d}_{x'}\varphi)$ at the point in question. By (2.6) the decaying phase has $\partial_{x^n}\phi = \mathrm{i}\sqrt{|\xi'|_g^2 - \tau^2/c^2} = \mathrm{i}\kappa$ there. Thus $\Im\phi = \kappa x^n + O((x^n)^2)$ along the inward normal. The factor $e^{-\lambda\Im\phi}$ gives the same leading penetration depth near a strictly elliptic boundary covector.

Where do we solve the equations in this construction? The exponential in (2.7) is an exact solution of the flat constant-speed model. For smooth variable coefficients, we determine the normal Taylor coefficients of the complex phase at the boundary by differentiating the eikonal equation there. The root is simple since $\kappa > 0$, so each differentiation determines the next normal derivative. With the boundary amplitudes prescribed, the transport equations determine their normal Taylor coefficients in the same way. We then extend these jets smoothly into a collar. The equations fix only the jets at the boundary. Thus the eikonal and transport residuals are $O((x^n)^N)$ for every N. Since $\Im\phi \geq bx^n$ for some $b > 0$ in a sufficiently small collar, multiplication by $e^{-\lambda\Im\phi}$ makes these flat residuals $O(\lambda^{-M})$ for every M.

Why might an exact phase fail to exist? In two spatial dimensions take a time-independent phase in the Euclidean half-plane $y > 0$, with $\phi(x, 0) = f(x)$ real and $f'(x) > 0$. The eikonal equation is $\phi_x^2 + \phi_y^2 = 0$. The decaying root gives $\phi_y = \mathrm{i}\phi_x$, the Cauchy–Riemann equation. At $y = 0$ this determines every normal derivative, $\partial_y^k\phi(x, 0) = \mathrm{i}^k f^{(k)}(x)$. An exact solution, however, would be holomorphic in $x + \mathrm{i}y$. Since its boundary values are real, Schwarz reflection would extend it holomorphically across the interval, forcing f to be real analytic there. Thus smooth nonanalytic data can give all the required boundary jets even when the complex eikonal equation has no exact solution with those data in a collar.

Now consider the interface problem below. A wave incident from S_+ can be propagating on that side and evanescent on S_-. Indeed, the tangential covector is the same on both sides. By (5.12), $|\xi'| = |\tau| \sin\theta_\mathrm{I}/c_+$. If $c_- > c_+$ and $\sin\theta_\mathrm{I} > c_+/c_-$, the decay rate in the transmitted medium is therefore

$$\lambda\kappa_- = \lambda|\tau|\sqrt{\frac{\sin^2\theta_\mathrm{I}}{c_+^2} - \frac{1}{c_-^2}}.$$

Here the transmitted side is $x^n < 0$. The normal root is now $-\mathrm{i}\kappa_-$, giving the same decaying factor $e^{-\lambda\kappa_-|x^n|}$. Why can there be a transmitted field in total internal reflection? For this flat-interface mode, $\partial_{x^n} u_\mathrm{T} = \lambda\kappa_- u_\mathrm{T}$ while $\partial_t u_\mathrm{T} = \mathrm{i}\lambda\tau u_\mathrm{T}$. Their product $\partial_t u_\mathrm{T}\,\overline{\partial_{x^n} u_\mathrm{T}}$ is purely imaginary. The time-averaged normal energy flux, which is proportional to its real part, is zero. This is consistent with the unit modulus of the reflected coefficient obtained from (5.13).

3. The Cauchy problem on the lateral boundary

Suppose we know both u and $\partial_\nu u$ on a small piece of $\mathbf{R}_t \times \partial\Omega$. Can we recover the wave just inside Ω? This time we prescribe two traces on the same timelike surface. We do not prescribe initial data at $t = 0$. Let h be the Dirichlet trace and k the outward normal derivative. As before, $x^n > 0$ inside and $\partial_\nu = -\partial_{x^n}$ at the boundary.

Take the flat case with constant speed c first. Fourier transform in (t, x') and write $s = x^n$, $v(s) = \widehat{u}(\tau, \xi', s)$. For each tangential frequency we have the Cauchy problem

$$v'' + \big(\tau^2/c^2 - |\xi'|^2\big)v = 0, \qquad v(0) = \hat{h}, \quad v'(0) = -\hat{k}. \tag{3.1}$$

This ODE has a unique solution for every frequency. To recover u we must also take the inverse Fourier transform. What matters now is how the solution depends on the data at large frequencies.

The hyperbolic region. Set $q = \sqrt{\tau^2/c^2 - |\xi'|^2} > 0$ and $\sigma = \operatorname{sgn}(\tau)$. By (2.5) the outgoing term has normal factor $e^{-\mathrm{i}\sigma q s}$, so write

$$v(s) = H_{\text{out}} e^{-\mathrm{i}\sigma q s} + H_{\text{in}} e^{\mathrm{i}\sigma q s}.$$

The traces give $H_{\text{out}} + H_{\text{in}} = \hat{h}$ and $\mathrm{i}\sigma q(H_{\text{out}} - H_{\text{in}}) = \hat{k}$. Solving, we get

$$H_{\text{out}} = \frac{1}{2}\left(\hat{h} + \frac{\hat{k}}{\mathrm{i}\sigma q}\right), \qquad H_{\text{in}} = \frac{1}{2}\left(\hat{h} - \frac{\hat{k}}{\mathrm{i}\sigma q}\right). \tag{3.2}$$

At normal incidence $\sigma q = \tau/c$. Thus a purely outgoing wave has $\hat{k} = (\mathrm{i}\tau/c)\hat{h}$, exactly the Fourier transform of (1.6). On a closed cone inside the hyperbolic region, q is comparable to $|(\tau, \xi')|$. Division by q gains the one derivative lost in taking the normal trace. Both oscillatory factors have modulus one, so the two traces determine the two waves stably there. For variable coefficients the same calculation with $|\xi'|_g$ separates their leading amplitudes at the boundary. We then solve the eikonal and transport equations along the two transverse ray families.

The elliptic region. Now put $\kappa = \sqrt{|\xi'|^2 - \tau^2/c^2} > 0$. Solving (3.1) gives

$$v(s) = \tfrac{1}{2}(\hat{h} + \hat{k}/\kappa)e^{-\kappa s} + \tfrac{1}{2}(\hat{h} - \hat{k}/\kappa)e^{\kappa s}.$$

The bounded half-space solution requires $\hat{k} = \kappa\hat{h}$, in agreement with the outward derivative of the decaying wave in Section 2.3. For general Cauchy data the growing term is present. For example, take $\tau = 0$ and $|\xi'| = N$ in a single mode. Boundary values $\hat{h} = e^{-\sqrt{N}}$ and $\hat{k} = 0$ give $v(s) = e^{-\sqrt{N}}\cosh(Ns)$. The data tend to zero faster than any power of N, yet the solution grows without bound at every fixed $s > 0$. This is why the unrestricted lateral Cauchy problem has no Sobolev stability estimate. Even existence of a distributional continuation for arbitrary smooth data can fail. For a curved boundary, $\sqrt{|\xi'|_g^2 - \tau^2/c^2}$ is the principal symbol of the map from the Dirichlet trace to the outward derivative for the decaying branch.

Glancing. At a nonzero covector with $\tau^2/c^2 = |\xi'|^2$, the two normal roots meet at zero. To see what survives of (3.2), let us combine the two terms and then take the limit. We get

$$v(s) = \hat{h}\cos(qs) - \hat{k}\,\frac{\sin(qs)}{q} \;\longrightarrow\; \hat{h} - s\hat{k} \quad \text{as } q \to 0.$$

This is exactly the solution of (3.1) at glancing. The full flat solution has a limit even though its incoming and outgoing amplitudes cannot be separated by (3.2) at $q = 0$. Geometrically, the ray is tangent to the boundary. For a curved boundary or variable speed, the behavior near tangency depends on how the ray meets the boundary. The construction by two transverse phases used above no longer applies uniformly there. At a simple glancing point one can use the Fourier–Airy construction of Melrose and Taylor [10, Chapters 6–8] to describe the transition across glancing.

4. Reflection of waves

Suppose an incident wave reaches $\partial\Omega$ at (t_0, x_0) along a ray transverse to the boundary. We work near this point, in the boundary normal coordinates used in (2.6). Thus $x^n > 0$ in Ω and $\partial_\nu = -\partial_{x^n}$ on $\partial\Omega$. Write the incident and reflected waves in the form

$$u_j = e^{\mathrm{i}\lambda\phi_j}\big(a_j^{(0)} + \lambda^{-1}a_j^{(1)} + \cdots\big), \qquad j = \mathrm{I}, \mathrm{R},$$

where $\lambda \gg 1$. We are given the incident wave. How do we choose the reflected one so that their sum satisfies the boundary condition?

The reflected phase. To cancel the incident oscillation on the boundary, we take $\phi_\mathrm{R} = \phi_\mathrm{I}$ there. Call their common trace $\varphi(t, x')$ and set $(\tau, \xi') = (\partial_t\varphi, \mathrm{d}_{x'}\varphi)$. The time and tangential derivatives are now fixed for both waves. By (2.6), the remaining derivative satisfies

$$(\partial_{x^n}\phi_j)^2 = \tau^2/c^2 - |\xi'|_g^2 \quad \text{at } x^n = 0.$$

Transverse incidence puts us in the hyperbolic region, so the right-hand side is positive. Which root do we take? Along the incident ray x^n decreases as t increases. Since $\mathrm{d}x^n/\mathrm{d}t = -c^2\partial_{x^n}\phi_j/\tau$ at the boundary, the incident root has the sign of τ. For the reflected ray x^n must increase. Thus the Cauchy data for the reflected phase are

$$\phi_\mathrm{R}|_{x^n=0} = \varphi, \qquad \partial_{x^n}\phi_\mathrm{R}|_{x^n=0} = -\partial_{x^n}\phi_\mathrm{I}|_{x^n=0} = -\operatorname{sgn}(\tau)\sqrt{\tau^2/c^2 - |\xi'|_g^2}. \tag{4.1}$$

The chosen root is simple, so the method of characteristics gives ϕ_R locally on the interior side. At the point of reflection the tangential velocity is the same for both rays and the normal velocity changes sign. This is the law of equal angles of incidence and reflection, measured in the metric g.

Dirichlet boundary condition. Take the trace of $u_\mathrm{I} + u_\mathrm{R}$ and factor out $e^{\mathrm{i}\lambda\varphi}$. For a zero Dirichlet trace the leading coefficient must vanish, so

$$a_\mathrm{R}^{(0)} = -a_\mathrm{I}^{(0)} \quad \text{on } \partial\Omega. \tag{4.2}$$

Neumann boundary condition. Now apply ∂_ν before taking the trace. Differentiating an exponential brings down $\mathrm{i}\lambda\,\partial_\nu\phi_j$, whereas differentiating its amplitude gives a term of order one. The coefficient of $\mathrm{i}\lambda e^{\mathrm{i}\lambda\varphi}$ in the normal derivative is therefore

$$(\partial_\nu\phi_\mathrm{I})a_\mathrm{I}^{(0)} + (\partial_\nu\phi_\mathrm{R})a_\mathrm{R}^{(0)}.$$

By (4.1) the two normal derivatives are opposite and nonzero. Setting this coefficient equal to zero gives

$$a_\mathrm{R}^{(0)} = a_\mathrm{I}^{(0)} \quad \text{on } \partial\Omega. \tag{4.3}$$

This equality gives the leading boundary value. At the next order the normal derivatives of the leading amplitudes also enter the Neumann condition.

Once we have the boundary value in (4.2) or (4.3), we find $a_\mathrm{R}^{(0)}$ inside Ω by solving the first transport equation along the reflected rays. For $P = \partial_t^2 - c^2\Delta_g$ it reads

$$2\big((\partial_t\phi_\mathrm{R})\partial_t - c^2 g^{ij}(\partial_i\phi_\mathrm{R})\partial_j\big)a_\mathrm{R}^{(0)} + (P\phi_\mathrm{R})a_\mathrm{R}^{(0)} = 0.$$

As a check, take a flat boundary, constant speed and normal incidence with $\varphi = t\tau$. Then $\phi_\mathrm{I} = \tau(t + x^n/c)$ and $\phi_\mathrm{R} = \tau(t - x^n/c)$. The outward derivative of the incident exponential gives $-\mathrm{i}\lambda\tau/c$ times its trace, and that of the reflected exponential gives $+\mathrm{i}\lambda\tau/c$ times its trace, in agreement

with (1.6). Thus we recover the minus sign in the Dirichlet formula (1.8) and the plus sign in the Neumann formula (1.10).

5. The transmission problem

To set up the stage, let $S \subset \mathbf{R}^n$ be an oriented smooth hypersurface with sides S_+ and S_-, and let $c(x)$ be a sound speed smooth on both sides of S, up to S but having possibly different jumps at S from both sides. We want to find a solution of the acoustic equation with that speed satisfying the additional transmission conditions

$$u|_{S_+} = u|_{S_-}, \quad \partial_\nu u|_{S_+} = \partial_\nu u|_{S_-}, \tag{5.1}$$

where we use the same unit normal ν, directed towards S_+, on both sides. If we use the outward normal of each side instead, the two outward derivatives sum to zero. To see where (5.1) comes from, write the equation as $c^{-2}u_{tt} - \Delta u = 0$ in the weak sense. A finite-energy displacement in H^1_{loc} has a common trace across S. Integrating by parts on the two sides gives interface terms with opposite normals. For a piecewise smooth solution these cancel precisely when the derivatives along the common normal agree. The conserved energy is

$$E(\mathbf{u}(t)) = \frac{1}{2}\int \left(|\nabla_x u|^2 + c^{-2}|u_t|^2\right) \mathrm{d}x \tag{5.2}$$

for solutions with finite total energy and no flux at infinity.

5.1. The 1D case. Consider the 1D wave equation

$$u_{tt} - c^2(x)u_{xx} = 0, \quad x \in \mathbf{R} \tag{5.3}$$

on the whole line with sound speed

$$c(x) = \begin{cases} c_+ & \text{if } x > 0, \\ c_- & \text{if } x < 0, \end{cases} \tag{5.4}$$

where c_+ and c_- are two positive constants. The transmission conditions take the form

$$u|_{x=0+} = u|_{x=0-}, \quad u_x|_{x=0+} = u_x|_{x=0-}. \tag{5.5}$$

Let $u_{\mathrm{I}} = U_{\mathrm{I}}(t + x/c_+)$ be incident from $x > 0$, with U_{I} compactly supported. For sufficiently negative t the packet has yet to reach $x = 0$.

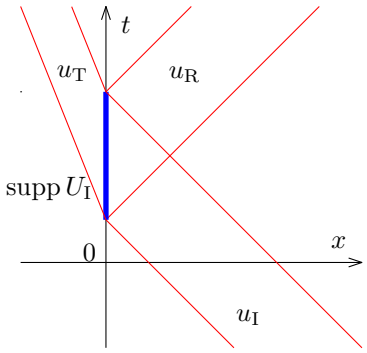


Figure 2. Reflection and transmission in time-space. Here, $c_- < c_+$.

We look for a solution of the kind

$$u = u_I + u_T + u_R, \tag{5.6}$$

where I, T, R stand for incoming/transmitted/reflected, respectively, and u_I and u_R are supported in $x > 0$, while u_T is supported in $x < 0$. Since we want the transmitted and the reflected waves to be solutions, moving away from $x = 0$, we must have

$$u_R = U_R(t - x/c_+), \quad u_T = U_T(t + x/c_-). \tag{5.7}$$

Then by (5.5),

$$\begin{aligned} U_I(t) + U_R(t) &= U_T(t), \\ \frac{1}{c_+}U_I'(t) - \frac{1}{c_+}U_R'(t) &= \frac{1}{c_-}U_T'(t). \end{aligned} \tag{5.8}$$

We view U_I as known. Adding the same constant to U_R and U_T preserves both equations in (5.8). We fix that constant by requiring the profiles to be compactly supported. Differentiate the first equation and use the second to solve for U_R' and U_T'. Integrating gives

$$U_R = \frac{c_+^{-1} - c_-^{-1}}{c_+^{-1} + c_-^{-1}}U_I, \quad U_T = \frac{2c_+^{-1}}{c_+^{-1} + c_-^{-1}}U_I. \tag{5.9}$$

If the speed is the same on the two sides, the whole incident wave is transmitted. For $c_- < c_+$, as in Figure 2, the reflected profile changes sign.

Energy. One can check directly that the sum of the energies of the reflected and the transmitted waves, once they have left the boundary, is equal to that of the incoming wave before it has hit the boundary. One way to do this is to notice that the coefficients in (5.9) satisfy

$$c_+^{-1}\left(\frac{c_+^{-1} - c_-^{-1}}{c_+^{-1} + c_-^{-1}}\right)^2 + c_-^{-1}\left(\frac{2c_+^{-1}}{c_+^{-1} + c_-^{-1}}\right)^2 = c_+^{-1}.$$

Then we use (5.2) to compute the energies above.

We can see the conservation directly from the transmission conditions as well. Differentiate the first equation in (5.8), multiply by the complex conjugate of the second, and take real parts. The cross terms on the left cancel, giving at each t

$$c_+^{-1}|U_I'(t)|^2 - c_+^{-1}|U_R'(t)|^2 = c_-^{-1}|U_T'(t)|^2. \tag{5.10}$$

To pass from flux to energy, notice that a packet $U(t\mp x/c)$ lying entirely in a medium of speed c has energy $c^{-1}\int |U'(s)|^2\,\mathrm{d}s$ by (5.2). Integrating (5.10) in t therefore gives the energy of the incident packet before it reaches the interface as the sum of the energies of the reflected and transmitted packets after they leave it.

5.2. Oblique incidence. What changes if the incident ray meets S at an angle? Take x^n to be signed distance to S, positive on S_+, so that the common normal derivative in (5.1) is $\partial_\nu = \partial_{x^n}$. As before, the incident wave comes from S_+. We work near a point of transverse incidence and write $c_\pm$ for the values of the two speeds there.

Write the three waves as $u_j = e^{\mathrm{i}\lambda\phi_j}(a_j^{(0)} + O(\lambda^{-1}))$, with $j = \mathrm{I}, \mathrm{R}, \mathrm{T}$. To match the oscillations in (5.1), take all three phases equal to the incident phase on S. If $\varphi(t, x')$ is this common trace, write $(\tau, \xi') = (\partial_t\varphi, \mathrm{d}_{x'}\varphi)$. These are derivatives of the phase. The full tangential covector of the oscillation is $\lambda(\tau, \xi')$ and is the same for all three waves. The eikonal equation on each side gives

$$(\partial_\nu\phi_j)^2 = \tau^2/c_\pm^2 - |\xi'|^2 \quad \text{on } S,$$

where $|\xi'|$ is taken in the induced metric on S. Suppose first that both right-hand sides are positive, and put $q_\pm = \sqrt{\tau^2/c_\pm^2 - |\xi'|^2}$. The incident and transmitted rays move towards decreasing x^n, while the reflected ray moves towards increasing x^n. Since the normal ray velocity is $-c_\pm^2\partial_\nu\phi_j/\tau$, we must take

$$\partial_\nu\phi_\mathrm{I} = \operatorname{sgn}(\tau)q_+, \quad \partial_\nu\phi_\mathrm{R} = -\operatorname{sgn}(\tau)q_+, \quad \partial_\nu\phi_\mathrm{T} = \operatorname{sgn}(\tau)q_-. \tag{5.11}$$

These roots and the boundary value φ determine the reflected and transmitted phases locally by the method of characteristics.

Let θ_I, θ_R and θ_T be the acute angles the rays make with the normal line, with angle zero at normal incidence. The spatial covector is parallel to the ray up to sign, so the sine of each angle is $|\xi'|/|\mathrm{d}_x\phi_j|$. By the eikonal equation $|\mathrm{d}_x\phi_j| = |\tau|/c_\pm$. Thus $\theta_\mathrm{R} = \theta_\mathrm{I}$, and we get Snell's law

$$\frac{\sin\theta_\mathrm{I}}{c_+} = \frac{|\xi'|}{|\tau|} = \frac{\sin\theta_\mathrm{T}}{c_-}. \tag{5.12}$$

The boundary system. Let $h_j = u_j|_S$. Given the incident trace h_I, we have two unknown functions on $\mathbf{R}_t \times S$, the reflected trace h_R and the transmitted trace h_T. Write $\Lambda_{\pm,\mathrm{out}}$ for the local outgoing Dirichlet-to-Neumann maps on $S_\pm$, and $\Lambda_{+,\mathrm{in}}$ for the incoming map on S_+. Each map uses the outward normal of its own side, as in (1.5). These derivatives are $-\partial_\nu$ on S_+ and $+\partial_\nu$ on S_-. Thus the two outward derivatives sum to zero in (5.1), and the transmission conditions give

$$\begin{pmatrix} \mathrm{Id} & -\mathrm{Id} \\ \Lambda_{+,\mathrm{out}} & \Lambda_{-,\mathrm{out}} \end{pmatrix} \begin{pmatrix} h_\mathrm{R} \\ h_\mathrm{T} \end{pmatrix} = - \begin{pmatrix} h_\mathrm{I} \\ \Lambda_{+,\mathrm{in}} h_\mathrm{I} \end{pmatrix}.$$

This is a 2×2 pseudodifferential system in the tangential variables (t, x'), microlocally in the region where both normal roots are simple and real. Indeed, taking the normal derivative of each local wave parametrix and restricting back to S gives a tangential operator of order one, modulo smoothing terms. Put $\epsilon = \operatorname{sgn}(\tau)$. By (5.11) the outgoing DN symbols are $\mathrm{i}\epsilon q_+$ and $\mathrm{i}\epsilon q_-$, while the incident one is $-\mathrm{i}\epsilon q_+$. Taking the principal symbol in each row gives the matrix

$$\begin{pmatrix} 1 & -1 \\ \mathrm{i}\epsilon q_+ & \mathrm{i}\epsilon q_- \end{pmatrix}.$$

Its determinant is $\mathrm{i}\epsilon(q_+ + q_-) \neq 0$. Thus the system is elliptic with row orders zero and one, and we can solve for both traces microlocally modulo smooth errors.

Let us apply this symbol matrix to the leading amplitudes $A_j = a_j^{(0)}|_S$. The first row gives $A_\mathrm{I} + A_\mathrm{R} = A_\mathrm{T}$. Dividing the second row by $\mathrm{i}\epsilon$ and moving the reflected term to the incident side gives

$$q_+(A_\mathrm{I} - A_\mathrm{R}) = q_- A_\mathrm{T}.$$

Substituting $A_\mathrm{T} = A_\mathrm{I} + A_\mathrm{R}$ gives the leading amplitudes at the interface,

$$A_\mathrm{R} = \frac{q_+ - q_-}{q_+ + q_-} A_\mathrm{I}, \qquad A_\mathrm{T} = \frac{2q_+}{q_+ + q_-} A_\mathrm{I}. \tag{5.13}$$

How much of the incident energy crosses S? For a real wave the energy density in (5.2) gives the flux $-u_t \partial_\nu u$ in the direction ν. Take the real part of $Ae^{\mathrm{i}\lambda\phi}$ and average over one cycle. At leading order the signed normal flux is $-\frac{1}{2}\lambda^2 \tau\, \partial_\nu \phi\, |A|^2$. All three waves have the same τ. The magnitude of the normal phase derivative, however, is q_+ for the incident and reflected waves and q_- for the transmitted wave. Dividing each outgoing flux magnitude by the incident one therefore gives the reflected and transmitted energy fractions

$$\begin{aligned} \mathcal{R} &= \left| \frac{A_\mathrm{R}}{A_\mathrm{I}} \right|^2 = \frac{(q_+ - q_-)^2}{(q_+ + q_-)^2}, \\ \mathcal{T} &= \frac{q_-}{q_+} \left| \frac{A_\mathrm{T}}{A_\mathrm{I}} \right|^2 = \frac{4q_+ q_-}{(q_+ + q_-)^2}. \end{aligned} \tag{5.14}$$

Adding the two fractions gives $\mathcal{R} + \mathcal{T} = 1$, as expected from conservation of energy. At normal incidence $q_+ = |\tau|/c_+$ and $q_- = |\tau|/c_-$. Thus (5.13) reduces to (5.9), and the flux weights agree with (5.10).

What happens as we approach the critical angle? Suppose $c_- > c_+$ and fix $\tau \neq 0$. By (5.12), as θ_I increases to $\theta_\mathrm{c} = \arcsin(c_+/c_-)$, the transmitted angle tends to $\pi/2$. Then $q_- = |\tau| \cos\theta_\mathrm{T}/c_-$ tends to zero while q_+ stays positive. By (5.13) the transmitted amplitude tends to $2A_\mathrm{I}$ for fixed incident amplitude, yet (5.14) gives $\mathcal{T} \to 0$. The transmitted ray becomes tangent to S, so the normal component of its energy flux vanishes in the limit.

The boundary system above the critical angle. Suppose now that

$$|\tau|/c_- < |\xi'| < |\tau|/c_+, \qquad \kappa = \sqrt{|\xi'|^2 - \tau^2/c_-^2}.$$

On S_+ the incident and reflected roots are still real. On S_- we take the decaying solution. To see which root to use, freeze the coefficients at the interface. The transmitted normal factor is $e^{\lambda\kappa x^n}$ since $x^n < 0$ on S_-. Thus $\partial_\nu \phi_\mathrm{T} = -\mathrm{i}\kappa$, and the outward derivative on this side is $+\partial_{x^n}$. Acting on the decaying factor it gives $+\lambda\kappa$ times its boundary value. The transmitted DN map therefore has principal symbol $+\kappa$ at (τ, ξ'). For a curved interface the phase and transmitted amplitudes are constructed as in Section 2.3, by solving the eikonal and transport equations to infinite order at S only.

In the 2×2 system for the traces we now use this decaying DN map in the lower right entry. The incident term on the right and the reflected term on the left are just as before. At the principal symbol level the system reads

$$\begin{pmatrix} 1 & -1 \\ \mathrm{i}\epsilon q_+ & \kappa \end{pmatrix} \begin{pmatrix} A_\mathrm{R} \\ A_\mathrm{T} \end{pmatrix} = \begin{pmatrix} -A_\mathrm{I} \\ \mathrm{i}\epsilon q_+ A_\mathrm{I} \end{pmatrix}. \tag{5.15}$$

Its determinant is $\kappa+\mathrm{i}\epsilon q_+$. The squared modulus is $\kappa^2+q_+^2=\tau^2(c_+^{-2}-c_-^{-2})$. Hence the determinant is elliptic of order one on every closed conic patch inside this region. We can still solve the boundary system microlocally for both traces, even though only the reflected wave propagates away from the interface. At the critical angle the limiting determinant is nonzero, but the transmitted square root loses smoothness there. The pseudodifferential argument applies away from glancing.

Solving (5.15) gives

$$A_{\mathrm{R}}=\frac{\mathrm{i}\epsilon q_+-\kappa}{\mathrm{i}\epsilon q_++\kappa}A_{\mathrm{I}},\qquad A_{\mathrm{T}}=A_{\mathrm{I}}+A_{\mathrm{R}}.$$

The numerator and denominator of the reflected coefficient have the same modulus. Thus the reflected wave carries all the incident normal energy flux, while the transmitted trace generates the evanescent field of Section 2.3.

Chapter VI

The Helmholtz equation in a domain with boundary

We study Helmholtz type of equations of the type $(-c^2\Delta - \lambda^2)u = 0$ in smooth domains Ω with boundary. If Ω is bounded, then $-c^2\Delta$ has a discrete spectrum and the boundary value problem is uniquely solvable only when λ^2 is not an eigenvalue; in fact the solution operator has poles at the eigenvalues. The relation between the solutions and the dynamical system consisting of geodesics reflecting off the boundary is delicate since it is affected by the long-time behavior of that system. Such an analysis is well behind the scope of those notes. In the hyperbolic case, we could just restrict the time interval over which we do the analysis, i.e., localize in time. We cannot do this here; in some sense λ is the dual of t and "feels" the time behavior for all t. For this reason, we will consider domains which are exterior to a compact obstacle. There can still be trapped trajectories which never escape to infinity. Our construction near the boundary is local in either case. In a connected exterior domain the outgoing solution at a fixed positive frequency is unique even when some rays are trapped. Where trapping makes a difference is in estimating the error of a parametrix as $\lambda \to +\infty$.

Roadmap

Our starting point is the outgoing wave $f(t - x/c)$ on the half-line in Chapter V. With the Fourier kernel $e^{\mathrm{i}\lambda t}$ of Chapter I and $\lambda > 0$, its spatial factor is $e^{\mathrm{i}\lambda x/c}$. The corresponding time factor is $e^{-\mathrm{i}\lambda t}$, so the full temporal Fourier covector is $-\lambda$. If we write the wave as $e^{\mathrm{i}\lambda\phi}$ with $\phi = \psi(x) - t$, the phase derivative is $\tau = \partial_t\phi = -1$ and the full tangential covector is $(-\lambda, \lambda\,\mathrm{d}_{x'}\psi)$. We will use (V.1.6) as a sign check.

The half-space problem. We will start with Dirichlet and Neumann data on the plane $x^n = 0$, with $x^n > 0$ in the domain. At normal incidence the outgoing Dirichlet-to-Neumann map is multiplication by $-\mathrm{i}\lambda/c$, and the incoming one by $+\mathrm{i}\lambda/c$, since $\partial_\nu = -\partial_{x^n}$. For general tangential frequency we will distinguish the two propagating roots, the decaying root in the elliptic region, and the failure of the simple-root construction at glancing.

A curved boundary. Next we will construct a local parametrix for oscillatory boundary data $e^{i\lambda\varphi}b$. We will use the inward distance coordinate of Chapter V, choose the outgoing normal root, and determine the boundary value for the first transport equation from either trace. Here $d_{\tan}\varphi$ is the tangential differential of the phase. The tangential frequency of the boundary oscillation is $\lambda\, d_{\tan}\varphi$.

Scattering by an obstacle. Given an incident wave, we will seek the scattered wave that makes the total Dirichlet or Neumann trace vanish. This will recover the reflection signs of Chapter V. How does a local choice of outgoing root fit the radiation condition at infinity? We will keep the normal pointing out of the exterior domain, into the obstacle. Far from it, where we assume $c = c_\infty$ is constant, the outgoing Sommerfeld condition on the scattered field selects the radial factor e^{ikr}, with $k = \lambda/c_\infty$.

From the parametrix to the exact solution. A small residual in the equation leads to a question about the resolvent. In a bounded domain we must specify the Dirichlet or Neumann realization of $-c^2\Delta$. Its resolvent has poles at the eigenvalues in the variable $z = \lambda^2$. In a connected exterior domain, outgoing uniqueness at a fixed positive frequency holds even in the presence of trapped rays. We will explain where a nontrapping hypothesis enters when we want estimates uniform at high frequency. This is where the long-time behavior of reflected rays matters.

We will close with a short transmission example. In the notation of the oblique-incidence calculation in Chapter V we take $(\tau, \xi') = (-1, d_{x'}\psi)$. The amplitude ratios are homogeneous of degree zero, while the DN symbols have degree one and pick up a factor of λ at the full covector. We can then read off Snell's law (V.5.12), the amplitudes (V.5.13), and the normal energy fractions (V.5.14) of that chapter.

Chapter VII

The elasticity system

Chapter VIII

Hyperbolic systems

Chapter IX

Weakly non-linear geometric optics

Bibliography

[1] M. Cekić, K. Krupchyk, S. K. Sahoo, and G. Uhlmann. The anisotropic Calderón problem for Riemannian metrics at high fixed frequency, 2026. Preprint, arXiv:2610.01280.

[2] M. F. Dahl. Geometrization of the leading term in acoustic Gaussian beams. *J. Nonlinear Math. Phys.*, 16(1):35–45, 2009.

[3] NIST Digital Library of Mathematical Functions. `https://dlmf.nist.gov/`, Release 1.2.8 of 2026-09-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.

[4] L. C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1998.

[5] L. Hörmander. Fourier integral operators. I. *Acta Math.*, 127(1-2):79–183, 1971.

[6] L. Hörmander. *The analysis of linear partial differential operators. I*, volume 256. Springer-Verlag, Berlin, 1983. Distribution theory and Fourier analysis.

[7] F. X. Kärtner. Fundamentals of photonics: Quantum electronics, 2006. Lecture notes, MIT OpenCourseWare, Spring term, available online.

[8] H. Liu, J. Ralston, O. Runborg, and N. M. Tanushev. Gaussian beam methods for the Helmholtz equation. *SIAM Journal on Applied Mathematics*, 74(3):771–793, 2014.

[9] S. Ma, S. K. Sahoo, and M. Salo. The anisotropic Calderón problem at large fixed frequency on manifolds with invertible ray transform. *J. London Math. Soc.*, 110(4):e13006, 2024. arXiv:2207.02623.

[10] R. B. Melrose and M. E. Taylor. Boundary problems for wave equations with grazing and gliding rays. Manuscript, available online.

[11] L. Oksanen, M. Salo, P. Stefanov, and G. Uhlmann. Inverse problems for real principal type operators. *Amer. J. Math.*, 146(1):161–240, 2024.

[12] J. Ralston. Gaussian Beams, 2001. Unpublished lecture notes, Institute for Pure and Applied Mathematics, April 18–19, 9 pp.

[13] S. Zelditch. Wave equation on a Riemannian manifold. Lecture 1, Summer Northwestern Analysis Program (SNAP), available online, 2017.