California Institute of Technology
Physics 2a Recitation
Nicholas Z. Rui

Setting the Record Straight
18 November 2022

"The only way for you to keep your mind straight is to run from those who would confuse you."
—Kristin Cashore

Contents

1 Why light goes in straight lines

These Notes will explain why light can usually be thought of as rays which travel in straight lines, even though it is actually a wave. These pictures seem totally inconsistent with each other, and it is not obvious to see why this may be done.

1.1 Diffraction patterns in the paraxial approximation

We will first consider diffraction from an aperture, with X being the distance along the optical axis and Y being the perpendicular distance (i.e., parallel to the screen). The setup is shown in Figure 1.

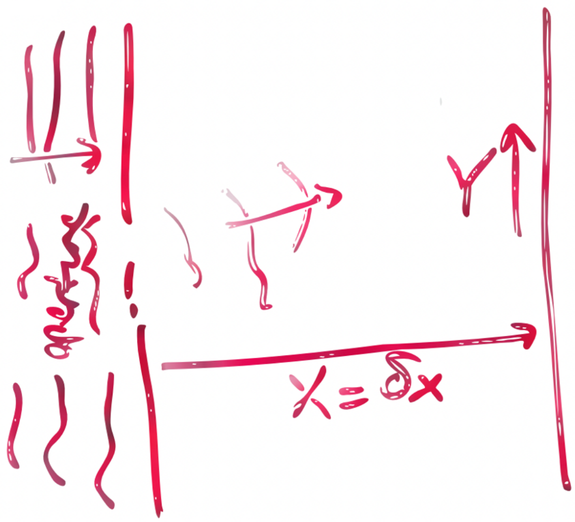

Figure 1: Setup of the problem.

According to the **Huygens principle**, every point on a wavefront emits as if it is itself a point source. In particular, this means that the field E a distance r away from a point source is

$$E \propto \frac{e^{ikr}}{r} \tag{1}$$

with the prefactor reflecting the magnitude of that field as well as the phase (which manifests as a complex prefactor).

A segment of the aperture at $(X, Y) = (0, y')$ and width $\delta y'$ produces a field δE on the screen, at $(X, Y) = (\delta x, y)$, which is given by

$$\delta E = E'(y') \frac{e^{ik\sqrt{x^2+(y-y')^2}}}{\sqrt{x^2+(y-y')^2}} \delta y' \tag{2}$$

When $\delta x \gg y, y'$, we can make the approximation that

$$r' = \sqrt{x^2+(y-y')^2} \approx x + \frac{1}{2}\frac{(y-y')^2}{x} \tag{3}$$

which is obtained by Taylor expansion. In addition, because the effect of the next-to-leading-order term is much more significant in the exponential, we can ignore it in the denominator. This is called the **paraxial approximation**, which is just a small angle approximation in optics where the distance to the screen is assumed to be much bigger than every other length scale.

Applying this approximation, we then have

$$E = \left[\frac{e^{ik\ x}}{x}\right] e^{ik(y\ \ y')^2=2\ x} E'(y')\ y' \tag{4}$$

where we have factored out the part of E which does not depend on y or y' (in the square brackets), i.e., is the same for every part of the aperture.

The function $E'(y')$ in Equation 2 is the field emanating from the aperture. Usually (but not always), we might be interested in the situation where the light source is itself a point source. If the point source is so far away that the phase across the aperture is uniform, then one can assume that f is a constant over the aperture and zero outside of the aperture, and we are said to be working in the **Fraunhofer regime**. Otherwise, if this phase difference is taken into account, we are working in the **Fresnel regime**. The difference between these regimes is shown in Figure 2.

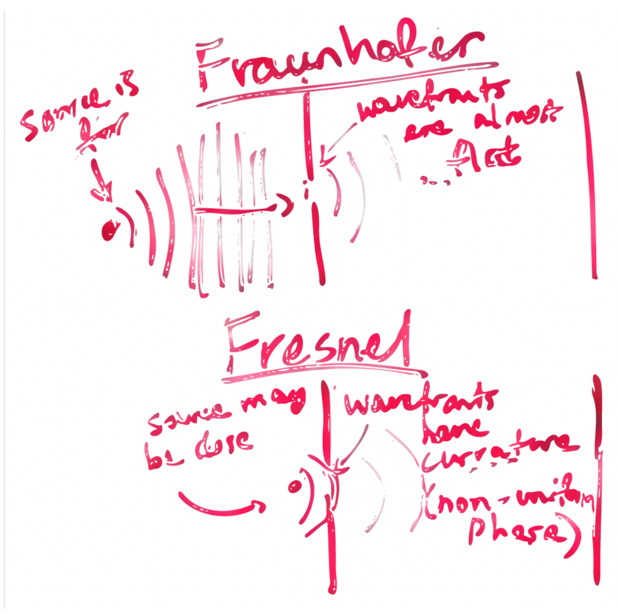

Figure 2: Difference between the Fraunhofer and Fresnel diffraction regimes.

However, the exact form of $E'(y')$ does not really matter here, except that it stores the amplitude and phase information of the wave coming out of the aperture.

The total field on the screen due to all of the aperture can be obtained by summing over Equation 4:

$$E = \frac{e^{ik\ x}}{x} \sum_{\text{all segments}} e^{ik(y\ \ y')^2=2\ x} E'(y')\ y' \tag{5}$$

which, when the segments of the aperture are taken to be very small, becomes an integral:

$$E(y) = \frac{e^{ik\Delta x}}{\Delta x}\int_{-\infty}^{+\infty} e^{ik(y-y')^2/2\Delta x} E'(y')\,\mathrm{d}y' \tag{6}$$

This can be squared to give the intensity.

1.2 The path integral of light in the paraxial approximation

We see that Equation 6 relates the field $E'(y')$ coming out of the aperture to the field on the screen $E(y)$. To make life more convenient later on, we will relabel $E'(y') \to E_0(y_0)$ and $E_1(y_1)$, i.e., Equation 6 becomes

$$E_1(y_1) = \frac{e^{ik\Delta x}}{\Delta x}\int_{-\infty}^{+\infty} e^{ik(y_1-y_0)^2/2\Delta x} E_0(y_0)\,\mathrm{d}y_0 \tag{7}$$

However, nobody said that there *has* to be a screen at $X = \Delta x$. If we instead remove the screen and place a new screen at $X = 2\Delta x$, then we can propagate the field at $X = \Delta x$ (i.e., $E_1(y_1)$) to $X = 2\Delta x$ (i.e., $E_2(y_2)$):

$$E_2(y_2) = \frac{e^{ik\Delta x}}{\Delta x}\int_{-\infty}^{+\infty} e^{ik(y_2-y_1)^2/2\Delta x} E_1(y_1)\,\mathrm{d}y_1 \tag{8}$$

As an aside, because one can "propagate" the field at X to $X + \Delta x$ by integrating over $e^{ik\Delta x}e^{ik(y-y')^2/2\Delta x}/\Delta x$, this combination of variables is called the ***propagator****.*

However, Equation 8 contains $E_1(y_1)$, which we have an expression for in Equation 7:

$$E_2(y_2) = \left(\frac{e^{ik\Delta x}}{\Delta x}\right)^2 \int_{-\infty}^{+\infty} e^{\frac{ik}{2\Delta x}\left[(y_2-y_1)^2+(y_1-y_0)^2\right]} E_0(y_0)\,\mathrm{d}y_1\,\mathrm{d}y_0 \tag{9}$$

However, we can then imagine removing *that* screen, and moving it to $X = 3\Delta x$. We would obtain analogously

$$E_3(y_3) = \left(\frac{e^{ik\Delta x}}{\Delta x}\right)^3 \int_{-\infty}^{+\infty} e^{\frac{ik}{2\Delta x}\left[(y_3-y_2)^2+(y_2-y_1)^2+(y_1-y_0)^2\right]} E_0(y_0)\,\mathrm{d}y_2\,\mathrm{d}y_1\,\mathrm{d}y_0 \tag{10}$$

If we repeated this process n times, we would have

$$E_n(y_n) = \left(\frac{e^{ik\Delta x}}{\Delta x}\right)^n \int_{-\infty}^{+\infty} e^{\frac{ik}{2\Delta x}\left[(y_n-y_{n-1})^2+\dots+(y_2-y_1)^2+(y_1-y_0)^2\right]} E_0(y_0)\,\mathrm{d}y_n\,\mathrm{d}y_{n-1}\cdots\mathrm{d}y_1\,\mathrm{d}y_0 \tag{11}$$

In other words, in this recursive process, we have obtained an (apparently unnecessarily) unwieldy formula for $E_n(y_n)$, which is a gigantic multidimensional integral where we must integrate over *every possible combination* of values of y that the light traveled through at every "screen" (i.e., every time X was a multiple of Δx). In other words, we integrated over **every possible path** from y_0 to y_n.

We can abbreviate this gigantic ugly integration as

$$\int_{-\infty}^{+\infty} \mathrm{d}y_n \, \mathrm{d}y_{n-1} \cdots \mathrm{d}y_1 \, \mathrm{d}y_0 \to \int \mathrm{d}[y(x)] \tag{12}$$

where we should imagine this integral as "summing over all possible paths" (and their associated phases) that the light could have taken. However, we should always remember that this is not a normal single-dimensional integral, but really the limit to an infinite number of nasty multidimensional integrals (which can, surprisingly, sometimes be done).

We can also investigate the exponent in the integrand:

$$\frac{ik}{2\Delta x}\left[(y_n - y_{n-1})^2 + \cdots + (y_2 - y_1)^2 + (y_1 - y_0)^2\right] = \frac{i}{2}k\sum_{k=1}^{n}\left(\frac{y_n - y_{n-1}}{\Delta x}\right)^2 \Delta x \tag{13}$$

However, for a large n, the light has traveled a total X distance of $n\Delta x \approx x$, so the sum in Equation 13 approaches the integral

$$\frac{i}{2}k\,\Delta x^2\sum_{k=1}^{n}\left(\frac{y_n - y_{n-1}}{\Delta x}\right)^2 \Delta x \to \frac{i}{2}k\int_0^x \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2 \mathrm{d}x' \tag{14}$$

Finally, Equation 11

$$E(y) = \text{const.} \int e^{\frac{i}{2}k\int_0^x \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2 \mathrm{d}x'} E_0(y_0)\,\mathrm{d}[y(x)] \tag{15}$$

where we are not overly concerned about the constant $(e^{ik\Delta x}/\Delta x)^n$, which does not depend on y or y'.

Equation 15 is called the **path integral**, and should be viewed as having the following interpretation: the field $E(y)$ can be obtained by considering the contribution of light coming from an aperture having taken every possible path, with each path contributing a phase ϕ, with

$$\phi = k\int_0^x \frac{1}{2}\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2 \mathrm{d}x' \tag{16}$$

The fact that each path carries a path-dependent phase means that paths can either interfere constructively (if the phases are the same mod 2π) or destructively, i.e., cancel out each other (if the phases are different by π mod 2π).

1.3 Paths of stationary phase

Before proceeding, we should first reflect on the phase in Equation 16. The prefactor includes the wavenumber, $k = 2\pi/\lambda$, which is usually very large because the wavelength of the light λ is, in most of our everyday experiences, extremely small. For example, for optical light, $\lambda \approx 0.5\ \mu$m.

Therefore, in most "big" situations, two closeby paths which deviate even slightly will have totally different phases, because the integral in Equation 16 is probably not so small that multiplying by

k won't make it large again. If the phase difference between two paths is significantly bigger than 2 (which is likely, as we argued above), this means that the phase difference between these paths will essentially be random, and adding up a lot of nearby paths will tend to cancel out.

There is only one situation in which the paths will not overwhelmingly cancel each other out, which is when a few paths which are nearby to each other have the same phase, to first order. For the phase in Equation 16, it can be shown that, if the initial y_0 and final y are fixed, the phase is *minimized* by straight line paths, i.e., dy/dx is constant. Therefore, even though Equation 15 is a sum over every possible path, *almost all of the paths are canceled out by their neighbors except for those paths where the phase doesn't change much*, i.e., $= 0$. This *could be* when is minimized (as in this case), although that was not required for this to happen. This concept is shown in Figure 3

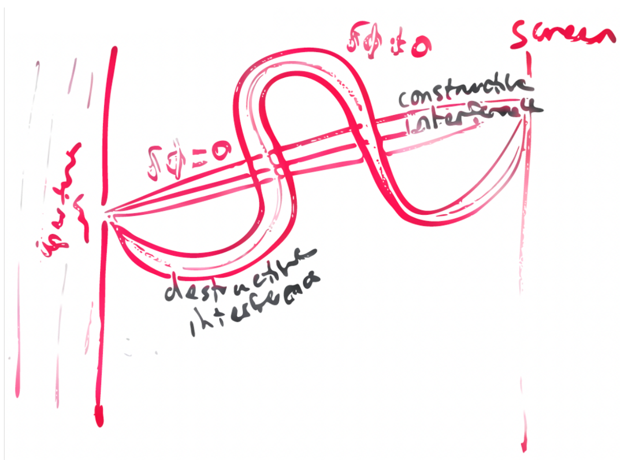

Figure 3: Paths of stationary phase (which constructively interfere) and paths of non-stationary phase (which destructively interfere).

The formal mathematics that goes into determining which paths in integrals like Equation 16 minimize, maximize, or "make stationary" those integrals is called **calculus of variations** (although we will not elaborate more here).

The assumption that *only* the stationary paths (with $= 0$) contribute to the field $E(y)$ is called the **stationary phase approximation**. A related law which arises from this is called **Fermat's principle**, which states that light travels in the distance which "makes stationary" (often, minimizes) its travel time. However, we see that, when the wavelength gets large, k is small, and so in this case other paths will cancel each other out less, and light will stop behaving exactly as it travels in just straight lines (i.e., its real wavelike nature will be apparent).

1.4 Other applications

Very strangely, in the eighteenth century, Joseph-Louis Lagrange realized that matter also seems to follow paths of stationary **action**, but in the sense of now requiring the following integral to be stationary:

$$\int_0^t L \, \mathrm{d}t' = \int_0^t (K \quad U) \, \mathrm{d}t' \tag{17}$$

where L (called the **Lagrangian**) is the difference between the kinetic and potential energies. This method, which is *entirely equivalent* to Newtonian mechanics, provided a powerful way to solve for the behavior of much more complicated systems.

Nevertheless, it was puzzling for hundreds of years why it happened to work out that a particle (which, in the Newtonian universe, only knows about the here and now) seems to behave as if it is checking over every possible path and choosing the one that fulfills a certain property.

With the advent of quantum mechanics, however, it became possible to write the wavefunction as a path integral in the same way (and actually, with more justification and fewer approximations):

$$(x; t) = \text{const.} \int e^{\frac{i}{\hbar} \int_0^t L \, \mathrm{d}t'} \quad (x_0; t_0) \, \mathrm{d}[x(t)] \tag{18}$$

where $\hbar$ is a very small physical constant, when compared with everyday scales. The analogy becomes even more clear when we think about the case where there is no potential. Then

$$L = K = \frac{1}{2} m \left(\frac{\mathrm{d}x}{\mathrm{d}t} \right)^2 \tag{19}$$

and

$$(x; t) = \text{const.} \int e^{\frac{im}{\hbar} \int_0^t \frac{1}{2} \left(\frac{\mathrm{d}x}{\mathrm{d}t} \right)^2 \mathrm{d}t'} \quad (x_0; t_0) \, \mathrm{d}[x(t)] \tag{20}$$

We see when looking for paths of stationary phase that those are exactly the ones that Lagrange realized particles must take in Newtonian mechanics. However, in quantum mechanics, when the difference between   between two paths is comparable to (and not much larger than) $\hbar$, the other paths no longer cancel each out, and the wavelike nature of particles can no longer be ignored.

It is interesting that hints of a quantum universe started to surface even hundreds of years before the advent of quantum mechanics, and the resolution to the puzzle that particles of matter seem to be capable of checking every single path even in classical mechanics.