

MITOCW | Lec 3 | MIT 2.71 Optics, Spring 2009

The following content is provided under a Creative Commons license. Your support will help MIT OpenCourseWare continue to offer high quality educational resources for free. To make a donation, or to view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at ocw.mit.edu.

GEORGE OK, so today I would like to cover-- to start, well, I will do a lot of things today. But to begin with, I would like to **BARBASTATHIS**: solve two applications of Fermat's principle. And the simplest possible thing that we try to do in optics, and that is to focus. So the one focus will come up again and again. It basically refers to the idea-- hi, Colin. It refers to the idea that we have rays, a bundle of rays that are propagating. And focusing means that we try to force these rays to meet at one point.

So this, of course, is very useful for imaging, but even simpler, if you want to collect all of the light energy, for example, for a solar concentrator, or a satellite dish, or something like that. Then basically, this is what you try to do. You try to get the rays that are coming from the source, and focus them into a point.

And this is valid, of course, in the entire regime of electromagnetic waves-- whether microwaves, or visible light, or even higher-- provided that the element that we use to focus the rays is much bigger than the wavelength. And here, I'm leaving it deliberately vague how much bigger. But think about it that is maybe a few thousand of wavelengths, or few hundred of wavelengths. In the microwave case, that's not exactly true. The RF wavelength maybe a couple of centimeters. The typical satellite dish is about the meter, so we're talking about 50 to 100 wavelength. But anyway, still the approximations are valid.

So the simplest way to do this kind of operation to focus the light into a point so it can be detected is with a reflector, a mirror whose surface is curved. And you know from experience if you look at the spoon that, well, funny things happen if you look at the spoon. We will talk about that perhaps a little bit later.

But you know that the spoon forms images, so this is basically a simplification or idealization of a spoon. So let's say that we have this surface. And for now, even though the label says parabolic reflector, I will pretend that I don't know the shape of the surface. So s of x is an unknown to be determined. And the goal, of course, is to have rays that are coming as parallel rays onto the parabol line, onto this unknown shape. And our goal is to focus these rays into a point on the axis of the surface.

And what you see. You see here that I put the symbol of infinity. I will mention that also a little bit later. If we have rays that are parallel, you can think-- according to Euclidean geometry, parallel rays never meet. Or another way to say it is that parallel rays, if they ever meet, that would be at infinity, at a very, very long distance away. So when the object of the imaging system is very far away, such as the sun, or a star, or a really remote source. Then we can say that the ray is coming from that object that are parallel, or that are coming from infinity.

So the objective here is, again, to determine the shape of the surface, s of x , such that the rays all focus at a point, f . So there's a number of ways of doing it. One is for example, I can apply the law of reflection anywhere in the surface. I have the array appearing here. I can compute the normal to the ray-- I'm sorry, the normal to the surface at that point of existence of the ray. And I can apply the law of reflection. I can solve the problem this way.

But that is actually a very complicated way of solving it. Much work. So being, I guess, lazy. The best way to solve this problem is to invoke the Fermat's principle. And Fermat's principle says, again, to remind you, says that rays that depart from a common point and arrive at another point.

They have to follow the minimum optical path. A corollary to that is that if you have two rays that start at the same point and meet again at the same point, then these rays must follow the same path. Because if one of those rays follows a shorter path, it means that the other ray that followed the longer path is violating Fermat's principle. And that cannot happen.

So the way we're thinking about this here is to take, for example, the center ray that is going directly on axis, that started at infinity. We go to the reflector. It will be reflected exactly backwards, and will go through the focal point, f . If we take another ray that is away from the central axis of the unknown surface. Then this ray also started at infinity. It gets reflected, and then, again, meets at that point, f . So therefore, the pattern that these two rays followed must be the same.

So how do we compute this path? Well, if you look at the central ray. The central ray is coming from infinity. I will use the plane passing through the focal point as my reference. And I will calculate the path back and forth from the focal point to the reflector and back. So I basically calculate this path. This path obviously equals 2 times f .

By f , I denote the focal distance of the reflector. That is the distance at which the rays come to a focus. And the other path I will compute is I will ignore the path of the ray up to the reference plane. And then I will compute this path over here. So how do you compute this path? Well, the shape of the paraboloid is s of x . So basically, s of x is the elevation of the unknown shape with respect to the other reference plane that passes through the bottom of the unknown shape.

So therefore, this distance here is f minus s , because it equals the focal distance minus the elevation of the reflector. And the other-- this distance over here. OK, so let me write it down. So it is f minus s . This is the first segment of the ray part. And the other is the hypotenuse of this orthogonal triangle, where one side of the triangle equals x . The other side of the triangle equals f plus s . And all I have to do is I have to apply Pythagoras' theorem.

OK, so this is the path of the ray that arrived away from the optical axis. And, of course, this must equal the path of the ray that is exactly on the axis that equals to $2f$. So now, we can play with his formula a little bit. I skip the derivation in the notes. Actually, I didn't quite skip the derivation of the notes. So instead of deriving it here, I will just animate it.

And if we do the algebra here, we can basically eliminate one of the focal distances, bring that to the other side. And then do a little bit more algebraic manipulation. And after not too much work, we have this expression over here for the elevation of the unknown shape. And clearly, what you see over that is a parabola. Of course, it turned out to be a parabola because I did the calculation in the cross section-- in one cross section of the optical element. In reality, all of these elements are rotationally symmetric surfaces of revolution.

So the reason I call it a paraboloid and not a parabola is because I take this shape, and then I spin it around the optical axis in order to get the familiar ball shape of the paraboloid this. But to do a little less math here, and to keep my life simple, I will only do these calculations in one plane. So turns out to be a simple parabola.

So then this is your final answer. This is the shape of a reflector that will take a set of parallel rays, and focus them into a focal point in front of the reflector. So these are very nice. And we actually got our first imaging system. We managed to compute its focal length as a function of the shape of the element.

There's a few things that I would like to remark here. One is, for example, that this works very nicely when the rays are indeed incident parallel to the normal at the center of the parabola here. If you imagine that I take this set of rays and I tilt it. Then a little bit of thought will convince yourself that it doesn't work so nicely anymore.

If you were to then come in parallel rays, then the rays will kind of focus, but they would fail to meet all at once at the focal point, f . So this is a perfect focusing element for on axis incident rays. But it would be not so perfect anymore if the rays are arriving off axis. That's one observation to keep in mind. This will come up again later.

One thing I would like to say is that this is a very nice element, indeed. But in some cases, it is not practical because as you can see, the focus is actually in front of the element. So if you were to detect the light, you would have to put, for example, a photo diode or some transducer at this location, which means you are actually blocking the path of the light.

AUDIENCE: I have a quick question.

GEORGE I'm sorry, I can not hear you.

BARBASTATHIS:

AUDIENCE: I can understand how the rays have a minimal optical path. But I don't understand why they have to have the same optical path. For example, you let it equal to $2f$, which was at the center distance. So isn't it possible that the minimum optical path does not equal $2f$? Or why is that the case?

GEORGE Take two rays. In our case, one ray [INAUDIBLE] from infinity, and they both meet at F . So the rays can take

BARBASTATHIS: different optical paths. One can go like this, the other can go like this. And doing our rays in general. So imagine that this has path I_1 , and this has path I_2 . And imagine, for the sake of argument, that I_1 is less than I_2 .

Fermat's principle says that if the ray, number one, goes through the first point from the origin to the destination. It follows a certain path. This path has to be minimal. But now, look at ray number two. This ray also goes from the same origin to the same destination, but follows a longer optical path. Therefore, the second ray has to violate Fermat's principle. So this is what we call mathematics abduction to-- what do you call it? To absurdum. That's right, abduction to absurdum.

In other words, by assuming that the rays have unequal optical paths, we arrived at a conclusion that is inconsistent with Fermat's principle. So therefore, this has to be discarded. The only possibility is that I_1 equals I_2 . So because the two rays start at the same origin and arrive at the same destination, they must have the same optical path. Did that make sense?

AUDIENCE: For example, in the homework problem, where you have the swimmer and the lifeguard, well, I guess-- I'll think about it a little more. Yeah, makes sense. Thanks.

GEORGE OK. So for the cases where the paraboloid is not convenient. Actually, I should say for the cases where the

BARBASTATHIS: reflector is not convenient, because of the problem of blocking the ray path at the focal point. It is perhaps more convenient to attempt to use a refractive optical element that is a dielectric. And try to focus the rays using now, again, a curved surface. But in this case, it is a curved surface refractive index n .

And for simplicity, I will assume that the rays are, again, are coming from free space, from vacuum, or from air. So the index of refraction is actually one on the left hand side over here. On the right hand side, it is n . So in this case, again, I have the same problem. I have an unknown shape, s of x , that I should follow, that I must somehow shape my refractor.

And one could assume that this is anything. It could be, again, a parabola. It could be a circle, I mean, a sphere. It could be a number of different things. The question, what is it really. So the question is actually the same as the question that I asked before. And to answer it, we will actually follow once again the same argument. We will use Fermat to argue that since the two rays started at the same point, infinity. And they meet at the same point, F , at the same target, F . They must follow the same optical path.

So in this case, I will apply Fermat with the reference plane that is tangential at the apex of the known refractive- - I'm sorry, of the unknown refractor shape. And I will compute the path of the own axis ray, which is very easy. So this path, you have to be a little bit careful now when we define the optical path. If you remember, the optical path in general. Optical path length equals the integral along the ray trajectory of the index of refraction times dl .

So what this means now in our case is that the optical path length for this segment of the rays between the apex and the focal point equals f , the distance, times the index of refraction, in order to convert it to proper optical path. OK, so that's easy. Now, we have to compute the other one. So here's, again, I will exaggerate it a little bit. Here is, again, my unknown shape, s of x . And here is the ray that is arriving at some distance, x , from the apex.

And this ray, according to my postulate here, will meet the axis at distance small f . To reach the point, uppercase f , that is the focal point. So again, I have to do a little bit of geometry here. This distance is s of x . Because this distance, again, is the elevation of the refractor shape relative to my reference plane over here.

So this equals s plus this distance, which again, is the hypotenuse of a triangle. But I have to be, again, a little bit careful because this segment over here now, again, is inside the refractor. So before I do anything else, again, I have to multiply by the index of refraction. So it has to be n , the index of refraction, times the hypotenuse length, which equals the square root of the two sides, of the two sides squared. What am I doing? I'm sorry. x squared plus f minus s squared. That's the correct expression.

OK, and [INAUDIBLE] in fact I could see it, because my computer has a preview. So I can see the equation, but I'm trying not to see it when I'm deriving things here. OK, so that is the equation, then. So again, according to the argument. Actually, let me repeat it once again, because this is a-- it's a very basic point.

We have two rays here. One is going on axis. There is no refraction here, because it is incident exactly normally on the surface. And then it propagates at distance, f . So the optical path length of this ray equals n times f . Then I have another ray that propagated a little bit further in there because it arrives at the elevated portion of the refractor. And then it bends in some way that we don't know yet. But what we do know is that we want, we demand, we require, so to speak, this ray to go through the same point, F .

So the fact that the ray goes through the same point, F. It means that it must follow the same path as the one axis ray. If this ray, for example, had followed a shorter path. Then there's no reason for this ray to go this way. It should have gone this way as well, because if it hadn't, it would be violating Fermat's principle. That is the argument. So therefore, there are two paths. This path, and this path. They must be exactly equal. So this is what this equation says over that.

OK, this equation now. Does anybody know what this equation represents? If you have looked at the notes ahead of time, you know the answer. If you ever looked at the notes, this equation doesn't mean much. However, it can be manipulated. And I did not do all the algebra in the notes. I skipped a few steps.

So the first thing you do is you basically try to get rid of the square root. So you bring this s on that side. You take squares of both sides. You do a little bit of algebra. You bring it to this form. Then this-- it still doesn't look very nice. So what you do is you try to write it to complete the sum of squares here. So that's a little bit messy. I will let you do it by yourselves, if you're interested and curious. If you get stuck, and you don't know how to do it, I will post the complete derivation on the website.

But anyway, after a little bit more algebra, you get a result that looks like this. It looks like a square of the surface shape, s , minus the displacement, times another square of the coordinate x equals a constant. And this type of equation in general, an equation of the form s^2 over constant, plus x^2 over constant equals another constant, which might as well be one. This equation represents an ellipse.

It is still not quite like this. It is of the form s minus some constant square over constant plus x^2 over constant equals 1. That is still in ellipse, but a displaced ellipse, right? What this really means is that the ellipse is centered at this location. OK, this comes from the displacement term in the square. And the eccentricity of the ellipse-- that is, the size of the minor axis relative to the major axis-- is given by the inverse of this quantity over here. So it is $n - 1$ over $n + 1$.

What is interesting to note is that the focal point is to the right hand side of the center of the ellipse. And in fact, this happens to be the focal point of the ellipse. If you look at the definition of an ellipse in the form of pure geometry, ellipses are characterized by two focal points. And it turns out in this case, the focus of the rays is one of the focal points of the ellipse. And that, of course, is symmetrically located on the other side.

OK, so this is a bit of a mathematical trickery. But what I want to emphasize, which is actually kind of important to know, is that the focal point is actually past the center of the ellipse. That's about it, really. The result of this is sort of an ideal elliptical shape. That's something worth remembering. That if you have a ideal set of parallel rays coming in-- or as we will call it a little bit later, a plane wave-- and we wish to focus it at the final distance inside of the electric material. The best possible shape to achieve that is an elliptical shape.

Now, again, there's caveats. For example, if I tilt this incoming ray, then the focus will not be perfect out here. So this is the same situation as in the parabol line. It only works so nice and so perfectly if the incidence is exactly normal. Out of curiosity, can anybody imagine-- so if you compare this with a paraboloid, one obvious advantage is, of course, that I'm not blocking the light path anymore. What is a possible disadvantage of this kind of approach? There are several, actually. Tell me at least one. Yeah. Push the button.

AUDIENCE: The detector is inside the vector.

GEORGE That's one obvious disadvantage. If I wanted to really use this, I would have to put the detector here. And, of course, it doesn't have to be a full ellipse. I can always clip the ellipse over here. The rays have no way of knowing that this happened to them, so I can always stick the detector over here. So maybe I can overcome that problem. But it is a problem, yeah. What is the second problem that kind of is sticking out?

AUDIENCE: In the case of if you're trying to build a telescope, or something where you have to worry about the structure of your optical setup. You could think that maybe a solid element would be much more bulky or hard to abrogate than a reflecting element, like paraboloid?

GEORGE That's correct. If you tried to make a big element, a big lens, a big focusing concentrator, like a solid concentrate, or a satellite dish. That's a very inconvenient shape to have. It is heavy-- absolutely, I agree. Tell me a third disadvantage. The paraboloid can go on a very long distance. This one is a finite shape because it's an ellipse. As it curves at some point, it will reach a derivative equal to zero. It will stop there.

So basically, this element captures only a finite set of rays up to the minor axis of the ellipse. It is missing the rest. Tell me a fourth and final disadvantage of this setup that at least I could think of. I'll give you a hint. Last time, Piper did a very nice demonstration with a prism. Do you remember something that his demonstration-- something about this demonstration [INAUDIBLE].

AUDIENCE: There can be total internal reflection. Like on this

GEORGE [INAUDIBLE].

BARBASTATHIS:

AUDIENCE: [INAUDIBLE]

GEORGE Is there a possibility for PIR here?

BARBASTATHIS:

AUDIENCE: The index of refraction isn't necessarily the same for all wavelengths?

GEORGE So first of all, let's answer the previous question. Here, you're right to think about the possibility of PIR. But if you look at it in a little bit more detail, you will see that PIR is not a problem in this case. At least for the rays going up to here, there's no problem of PIR. Because the rays come from a low index medium to a high index medium. And therefore, they will always refract inside the [INAUDIBLE].

What is interesting to happen to the ray that is incident exactly tangential to the ellipse over here. It will still be refracted, if you think about it. It's Snell's law. And it will give you sort of the maximum acceptance cone of the interface that we discussed last week, I mean, on Monday in relation to Snell's law. So then out of the two demos that Piper did last week, the PIR is not the problem. It is the other one, as you correctly pointed out. So you want to repeat? Who has spoken before? Someone mentioned the index of refraction.

AUDIENCE: The index of refraction isn't the same for all wavelengths.

GEORGE That's right. That's right. So since the index of refraction is not the same for all wavelengths, what this means is that the focal distance now becomes a function of the index of refraction. And I may tune it. Basically, what I do is I manufacture this shape. So I can manufacture it for a certain index for a standard wavelength. If the index changes, this relation will not be satisfied anymore, and therefore the refractive element will become imperfect.

How do we call this phenomenon of dependence of index of refraction on wavelength? Dispersion. Dispersion. And in the context that we're discussing here. OK, it is dispersion. In the context of focusing, when we get imperfect focusing because of dispersion, it has also another special name. It is called chromatic aberration.

The term aberration we'll see again in the next hour. It generally refers to failures of optical systems to focus light as they're supposed to. So this is the first such failure that we encounter. The term chromatic, I have an advantage over you. I'm Greek. And this turns out to be a Greek word.

Chroma in Greek means color. And this is the Greek spelling. If you want to spell it out in Latin so you can pronounce it, it is chroma. And that's where the word chromatic is derived from. It literally means separation due to color. Any questions on any of this so far? I'm sorry, there's a question.

AUDIENCE: If someone thought of a lenslike element in which the focusing takes place on the other side of the element. Will it have a [INAUDIBLE] power surface?

GEORGE That's coming up, actually, in a few slides. So yeah, I mean, the answer to your question is generally, if you want
BARBASTATHIS: to focus a point from the left to a point on the right, at the opposite side of element. You need two hyperbolic surface. I'll cover that in a second. Of course, hyperbolas are difficult to manufacture. People typically make spherical surfaces.

You can also have a sphere, nowadays, by injection molding. I will go into that also later. But even that is only perfect for on axis points. If you move points off axis, then it fails again. So it becomes the whole problem of optical design. OK, so I will switch now to today's, actually, set of lectures.

And you realized that we were kind of behind by almost one lecture. I think we're beginning to catch up now. But what I would like to do today for the rest of the two hours that you have is to go relatively slowly over a number of different topics, which are also related to focusing. But as we go along, we will generalize focusing to more and more complicated situations.

So I wanted to solve two cases of-- actually, one case of focusing using two different types of elements, a reflector and a refractor. In order to make this stuff a little bit more specific, let me remind you of some definitions that we discussed a little bit at the first lecture. But I would like to remind you again with a bit more detail.

So what is the concept of a spherical wave. So spherical wave is what you get if you have light or originating at the point source. It is also referred to as a point object. So what happens then is you get the fan of rays that is divergent. And if you plug the normals to these rays, these normals, they look like spheres. So, of course, the wavefronts expand as the spherical wave propagates. And because these wavefronts are spherical surfaces, that's why we call it a spherical wave.

Again, very important to remember by definition. The wavefront is normal to the rays. So what I have done here in principle is I took each ray. I computed the normal to the ray. I connected all the normals, and I get this sphere over here. I don't know if I did it very well in my graphic over here, but that's what this is supposed to denote.

OK, you can flip this situation and create a convergent spherical wave, where basically, you have rays that are propagating towards the point. If that is the case, it is not a point source anymore. You call it a point image. In most optical systems-- at least in optical systems that we will consider here-- this is really the goal of an optical system.

You try to take rays of light that are given to you by some source that you do not control, such as an object, or a star far out of reach. And what you try to do is you try to arrange your optics, select your lenses, space them appropriately, and so on and so forth. So that you can converge these incoming rays into point images.

So if you wish this is our given, or our customer. And then this is what the customer wants. The customer wants to form point images. So a good deal of working the field of optics has to do with this problem, how you can image-- however you can create point images. I'll take this out, because it does not relate to this discussion. OK.

So again, the wavefronts are spherical here with a common center at the point image. But in this case, of course, the wavefronts are contracting as you go towards the point image, because the rays are converging. And, of course, [AUDIO OUT] wavefronts ideally collapse into a point. This doesn't sound very physical.

And a little bit later, when we do the theory of diffraction. We will see in a bit more detail what happens at that point. It is not really that the wavefront ideally collapses. Something else happens that we will see. Within the geometrical optics approximation, let's accept it for now that this is the case. That indeed, the rays can converge to a single point, and then the wavefront over there collapses.

OK, this is true for pretty much every case of imaging. However, in a number of situations, it is very convenient to take one extreme case where the point source, or the point object is very far away. I mentioned that already about half an hour ago. And I called it the point source of infinity, or a point object of infinity. A very good approximation for that is a star.

If we look at stars in the night sky, they look like points. Of course, a star is humongous. It is probably thousands of times bigger than the earth. But because it is at a distance of thousands or millions of light years away, it appears to us to be like a point, a point source. So again, there's another important point to emphasize.

When we talk about the point object at infinity, we don't necessarily imply that it is mathematically a point, that is, it has zero dimension. What we really mean is it is so far away that the angle that it subtends towards that far away point is minimal. If that is true, if that assumption is true, then I can call it a point. I mean, I can call it a point source at infinity.

And what happens if you really went near that point far at infinity, you would still see divergent rays. But because these rays propagate at a very long distance, you only get to see the set of rays near the axis that departed from that point. And because you're [INAUDIBLE], these rays are really tightly collapsed near the central ray over here. They all look parallel.

So a set of rays that I propagate and they're parallel. Like so on here, we call it a plane wave. If you go back to your notes from the first lecture, we called it the plane wave, or the planar wavefront.

But we can also call it-- parallel [INAUDIBLE] rays, we can also call it an object at infinity. Of course, because the rays are parallel. If I draw the normals to them, they will form planes. They form ideal planar surfaces. And that's why this is called the planar wavefront, or a plane wave. For our purposes, the two terms are identical.

And, of course, this goes the other way too. If I had a planar wave also propagating like this. I can also claim that I create a point image at infinity. Because if I take parallel lines and I propagate them for an infinitely long distance. OK, what I'm about to say is not mathematically correct, but in optics, we say it all the time. It is that these parallel rays will meet at infinity.

Of course, parallel rays never meet. But you can simply to justify it in your mind, flip to this picture, and take these rays, propagate them backwards. At a very, very far away distance, you cannot tell the difference, whether they started as parallel rays, or they started as a tightly convergent small ray bundle.

OK, so this [AUDIO OUT] because very often, it simplifies calculations and gives certain kinds of intuition. And again, we define it in the geometric optic sense. When we do diffraction about a month from now, we will define it more rigorously in a way that is probably much more convincing than I do now. The reason I'm apologizing for this definition is that there's no such thing as mathematical infinity in real life. Even a star, it is a very long distance away, but it is not an infinite distance away.

So what do I mean by infinity? For example, [INAUDIBLE], who is sitting here on the front row. Is he at infinity with respect to me, or no? Well, what about you guys who are sitting in the cam bridge? Well, OK, the surface of the earth is curved, so that creates an obvious problem. But suppose I could draw a straight line through the center of the earth from Singapore to Boston. Would you guys be at infinity with respect to me?

OK, it actually depends on a lot of different things. It depends on the relative sizes that I have and you have. It depends on the wavelength that the light is propagating. And actually, that's about it, according to diffraction theory. So we will see a rigorous definition of what is infinity when we do diffraction. For now, let's take it on faith that if I have a set of rays that appear [AUDIO OUT], we'll call it image of infinity.

So sometimes, we use the term collimated. That's a better marker. Collimated rays. So collimated is from the word colinear, right? Creating parallel to each other. OK. So I included this slide, but as a reminder, and that as a reference, because for the next-- I don't know, for the next semester really, from now on, we'll be using these terms a lot. Spherical waves, plane waves, spherical wavefronts, and so on and so forth.

The concept of infinity applies to both convergent and divergent spherical waves at infinity. All right. So the two examples that we saw at the beginning. These are basically examples of imaging sources at infinity. Because we had plane waves coming into the paraboloidal reflector and ellipsoidal refractor respectively. And then these two elements managed to convert these plane waves, these parallel ray bundles into perfect converging spherical waves.

Therefore, they created perfect point images at the corresponding focal points. And then wavefronts, I also put down the two equations, the equations of the two elements. Now, let's flip the coin a little bit, and ask how can I create a point image at infinity? Now, I start with a point source that is within my range. Here is my point source. Here is my point source. And I'm trying to create an image of that source of infinity.

Well, another way to say it is that what I'm trying to do is I'm trying to collimate the light. So in the case of the paraboloid, it's actually very simple. All I have to do is reverse the paths of the rays.

This may also sound arbitrary, but in electromagnetics, there's a very basic principle that says that you can do that. I will not go into that in detail, but I can assure you that from the rigorous electromagnetic point of view, this is permissible. You can reverse the ray paths, and you still get the value of the electromagnetic field.

Therefore, a paraboloidal reflector works both ways. You can use it either as a receiver, so you can focus a source at infinity, and you can detect it here. Or it's a transmitter. You can have a point source at this finite distance, and you can broadcast a number of cars, but you can beam it. That's the word. You can beam it to a very narrow-- to a target that is at a very long distance away.

In the case of a refractor, the answer is not as simple. Because in the process of reversing the situation, now, the source is in air. And the collimated light is in glass, is in the dielectric. So therefore, the ellipse doesn't necessarily provide the same answer anymore. So basically, you have to solve the problem again. And indeed, we find that the ellipse is not the answer in this case. The answer is a hyperbola. A hyperboloidal surface.

Given by this equation over here. If you compare it with the case of the ellipse, the difference is a minus sign. So again, I will not do this derivation. I will let you do it by yourselves if you have the inclination. Or if we have time at the end today, we might do it. It's a very similar derivation.

We apply Fermat's principle. We basically demand that the path from here to some reference point equals the path from here to the same reference plane off axis. When you apply this principle, you end up with this equation after some algebraic manipulation.

So the word confusion. I actually put together a table that contains all the possible cases of first of all, reflective versus refractive focusing elements. And point sources, or point images at infinity. And there's actually six possible cases.

For the kinds of a reflector, we saw that it doesn't really make a difference. In both cases, a paraboloid is the answer. Because I computed already the focusing of an object at infinity, which is focused at the focal point of the paraboloid. And if I reverse the rays, the same paraboloid will take a point source at F, and image that point source at infinity. So that is very simple.

In the case of a dielectric, I can do two things. First of all, I start with the ellipsoid. When we derive the case of the ellipsoid, focusing a plane wave coming from infinity. And we saw that it focuses at the focal point of the ellipse, inside the ellipse.

If I reverse the ray paths, then I have this situation where the light originates at the center of the ellipse. And then it gets collimated as it comes out. Now, the reason the ellipse is still the answer here is because the light starts at the dielectric. So really, now, I'm correct. I can still revise the optical path, and I get the correct answer.

So this has led to cases of objects at infinity, and image at infinity. But have to be careful that the object is in air. In this case, the object of infinity is in air. The focus is inside the dielectric. In this case, the source of the object is inside the dielectric, and the image at infinity is in air.

The other case is the one that I just mentioned, and did not derive, but I let you derive by yourselves. And that is the case of a source that is now at a finite distance. And I wish the image to be at infinity inside the dielectric. So this case, this is hyperboloidal surface.

And, of course, you might ask, well, what the heck does it really mean? Can I really make an infinitely large refractor? Well, that says you don't have to, because if any of the refractor is cut here. Since the rays are coming out normal, they do not refract at all at this interface. So really, you can chop the hyperbola, and then you can still have your point, your image at infinity.

So that's fine. And, of course, you can also reverse the ray paths here. And you can create the opposite situation where whenever you have an object at infinity, but inside the dielectric. And you have a point image at a finite distance in air. And that could also be hyperbolic, right? Because this, I derive simply by flipping the ray paths from one case to the next. OK, any questions about this?

AUDIENCE: I've got a question. If your point sources within a green lens, what is the corresponding index of refraction function?

GEORGE Within a green lens?

BARBASTATHIS:

AUDIENCE: Yeah.

GEORGE I will cover green lenses in about two weeks. We will do Hamiltonian optics, so yeah.

BARBASTATHIS:

AUDIENCE: I also have a question. So if you said with the hyperboloidal refractor, if you cut off the back end, the rays will continue to come out parallel. Does that work in the opposite direction? Like if you have a lens that's flat on one end and a hyperbola on the other, can you focus it to a random point? Or not random.

GEORGE Yes. This case, for example, that would work.

BARBASTATHIS:

AUDIENCE: Yeah, I can't see.

GEORGE If you have rays coming in, yeah.

BARBASTATHIS:

AUDIENCE: [INAUDIBLE] right, yeah.

GEORGE Of course, the problem if you do it this way is that you have to cut off the hyperbol at some point. So you cannot

BARBASTATHIS: extend your incoming bundle infinitely.

AUDIENCE: Thanks.

GEORGE Of course, these are all very highly idealized situations, I should emphasize. Any other questions? OK, so now

BARBASTATHIS: that you have seen these, I would like to say a few things about wavefronts, and what happens to the wavefronts as they go through of these--

AUDIENCE: George, I think that's another question.

GEORGE Oh yeah, I'm sorry. Yes.

BARBASTATHIS:

AUDIENCE: So would the ellipsoidal reflector also be more subject to spherical aberrations since the focal point is actually within the optic instead of in front of it?

GEORGE Say that again? If it is in front?

BARBASTATHIS:

AUDIENCE: Yeah, it's comparing the--

GEORGE OK, I have not yet defined what is spherical aberration. Assuming that we know what spherical aberration is. All

BARBASTATHIS: of these are free of spherical aberration. By construction, they focus light perfectly. However, a spherical aberration will come in, for example, if you-- let's take this case. Yeah, if you move the source off axis, then the plane wave of coming out will be aberrated. The same here.

If you tilt this plane wave. Again, the point image will be aberrated. Actually, it would be aberrated in more than one way. It will contain coma as well as spherical. So these are free of aberrations to the degree that they are also operated as design. But we'll say a little bit about aberrations in the coming slides. So maybe you can postpone that. Any other questions?

So speaking of aberrations, here they come. First of all, let me say a few things about the wavefronts. So by definition, I will just look at this case, and then you can construct similar diagrams for all the six cases that were in the previous slide. I didn't want to do that. It actually takes a very long time to do such a slide, so I didn't want to do all of them. But this is pretty representative. So it conveys the idea, I think.

So by definition, we have a plane wave coming in. So you have plane wavefronts arriving at the interface. And also by definition, because by construction we required all the rays, all the refractive rays. They are required to meet at the focal point over here. It means that what we have inside the refractor is converging spherical waves. And therefore, if you look at the wavefronts collapsing in towards the focal point, they are perfect spheres.

This is all equal section, so these are perfect spheres collapsing towards uppercase F. It's kind of interesting to think what happens here at the interface. So [INAUDIBLE] the interface, you have two things happening. You have the ray arriving at the left, then sort of being refracted into the sphere. But the wavelength has to remain-- well, it has to remain continuous.

So basically, what happens is the wavefronts, they curve in a continuous fashion, as shown here. You can think of it basically as the wave approaches from the left toward the refractor, as it starts entering the refractor, of course, it enters first at the apex. And then the wavefront starts bulging. It sort of creates this bulge. And then as it goes more and more in, the bulge progresses to create a segment of a spherical surface.

Now, you can justify that with a number of different ways. One is, of course, Fermat's principle that we just applied. The other is if you think that light actually propagates slower in this medium than it propagates in air, then you can see why the wavefront starts bulging here. Because as soon as the light enters the medium, it gets slowed down.

So now that portion of the wavefront that is still outside in air is actually faster than the portion that is here. Even if you compare, let's take this case. If you compare the portion of the wavefront that is here at the apex to the portion that is here, to the portion that is here. You can see that this one has traveled the longer distance in the medium.

So that's why it is further behind. This one has traveled slightly less, so therefore, it is a little bit further ahead. And this one is still in air, so it is even more ahead. So this is another way to explain why the wavefronts are bulging, and they become spherical as the wave enters the medium.

Now, of course, they become exactly spherical. It's very special to shape that we chose. Because we demanded Fermat to hold, the wavefronts that result from refraction of this elliptical interface. They actually become spherical. But you can easily convince yourself that if I chose a different surface.

Say, if I chose a sphere or a hyperbola in this case, or some crazy polynomial, quartic, sixth order, I don't know, some generalized surface. Then the wavefronts that we get here. They might still bulge, because the light would enter a slower middle, but they wouldn't be spherical.

Now, the fact that the wavefronts would not be spherical. What it really implies is that the focus is not perfect anymore. If these surfaces of the wavefronts, they deviate from spheres. Then the rays would not really all be at $[F, ?]$ but they would cross at various places. And that is what you call an aberrated image, or an aberrated wavefront.

So the two are related, because if, for example, the wavefront deviates from a sphere, then you call it an aberrated wavefront. If the image deviates from the ideal point image, where all the rays meet to a sort of blurry image where the rays fail to meet, then that's an aberrated image. So I will show some examples of that later on, today, and also later during the class.

So there's a special case when the-- well, let me save that special case for later. I will talk about this in a little bit. What I would like to say next is why is it so significant to-- why do we try so hard to create these point images?

We're surrounded by objects that are-- actually, most of the objects that we see around us are reflective. They are opaque. So light that is entered in from a light source, for example, a light bulb, or the sunlight when we're outdoors.

What happens to the light as it hits the objects that are surrounding us. It scatters. So scattering really means that the light that is arriving from a source. Each point in the object, it creates a highly divergent spherical wave.

So basically, each point in the object becomes a secondary point source. This is part of a big-- of another principle in light propagation called the Huygen's principle. But we're not quite ready to see that principle yet. It needs a little bit more careful definition.

So for now, one way to justify in your mind why this might happen is that the objects surrounding us. They are composed of atoms. And each atom, when it gets hit by the light beam. It will be set in motion. And you can imagine the situation where the motion of the atom will irradiate a new light beam. So these light beams that are coming out of there of the atoms composing the surfaces surrounding us. This is what we perceive as-- I should say, what we receive as scattered light.

And it sounds a little bit like hocus pocus, but it's actually not very far from reality. You can justify the argument I just made with some very basic principles. So it is more or less correct. The bottom line is that the illumination becomes these divergent spherical wavefronts.

Now, if you are sitting here, and you're observing all the rays coming from the object. You can imagine that it's a mess. For example, here, you have rays coming from this point, and from this point, and from this point, from everywhere, everywhere inside the object. You have rays arriving here.

So it is very difficult to understand what's going on. All of these rays, they create kind of a very blurred image, very similar to what happens to those of us who have glasses. When we take off our glasses, everything appears blurry. Well, it is not as bad as it would have been in this case. But the reason things are blurry in our eyes is because when we take our glasses off, we cannot form that effect-- well, it's never perfect, but we cannot form good enough quality images in our eye.

So what we need in order to design these kind of situations is an optical system, or a sequence of optical elements that will pick up these divergent spherical wavefronts, and will convert them into converging spherical wavefronts. And if we can somehow create an optical system that will take each and every one of these wavefronts.

Now, we're talking, of course, about an infinitely large number of point scatterers that are generated here. But if you can somehow take each and every one of them, and I can convert it into a converging spherical wave, then when I put all these converging spherical waves together, I will actually get an image, which will be as close to perfect as I could possibly want within the limits of approximation of geometrical optics.

Another way to express it, which is really like a fancy mathematical term, but it is quite appropriate here is the imaging system creates a map. Creates a map from point sources in the object space to point images in the image space. And this map is, of course, very desirable.

Because if it fails, if I tried to detect the image over here that is before the rays have come to focus. Then the map becomes ill-defined, because each point over here is receiving arrays from multiple points in the object. This is exactly the same situation I was referring to before, when we take off our glasses. Those of us unfortunate enough to require glasses.

For the rest of you who do not need glasses in everyday life, you can, of course, create that same situation by putting on a pair of glasses that you can borrow from one of your colleagues. So in both cases, you can add what we call the focus, which basically means that the image is not just imperfect, but it is blurry. It is very far from perfect.

However, from what you must have gathered already from the previous discussion, this is a very difficult task. In fact, it is an impossible task. Because you saw that even in the case of a simple, ellipsoidal, or hyperboloidal focusing element. You could see that the one to one mapping works only on axis. If you tried to focus simultaneously points off axis, well, then, of course, it stops to work. It is very easy to convince yourself in geometrical terms.

Even then, there is some special cases where you can have an entire surface that is in focus. But we will talk about this later. So this is actually-- you could think of it as a very unfortunate event. I think it as a fortunate event, because it keeps people like me, and Professor Sheppard, and a lot of colleagues, it keeps us employed. Because this is the whole reason people need optical systems is they need people like us. So it is a very fortunate fact.

But what I'm trying to say here is that this job of perfect imaging cannot be done exactly. So therefore, the science, or if you wish, the engineering of designing optical systems is to arrive at the best compromise within the cost, constraints, the elements that you have available, the physics, of course, this is definitely inviolable. You can violate a budget, right? You can run a budget into the red, as Wall Street banks discovered in the last six months. But you can certainly not violate physics. Even bankers cannot violate physics.

So within all these constraints imposed by physics, economics, and so on so forth, how can you make an imaging system that does the best job possible? OK, so let's go back to this question of how can we create good enough quality images so that we can design imaging systems?

So let's look at a slightly different case, now. So far, we saw pairs of images and sources, where one of the tools at infinity. So objects at infinity, finite images. I mean, finite distant sources and images at infinity. We start to deal with this case, where we have a point object and a point image. So both are now at finite distances.

Based on what we saw so far, and assuming that you're not looking at your notes. The answer is already in your notes, if you reach that page in the notes. But if you haven't seen that, then your time to guess. What would be the ideal imager for this case that would take a point object and focus it perfectly into a point image on axis? Did we not answer already based on--

AUDIENCE: The concatenation of two hyperbolic reflectors.

GEORGE That's right. So that is not necessarily one way of doing it, but it's certainly one way of doing it based on what we
BARBASTATHIS: learn. If I put two refractive surfaces like this, I can arrange the first one to be a hyperbola. What it will do is it will collimate. Yeah, it will collimate the incident divergent wavefront. And then if I can arrange for the second one to be also hyperbola, then that will focus. The plane wave that was propagated here, it will focus into a perfect point image.

So this is now, again, a perfect imager. But now, it works for finite distances. One focal distance in front of the hyperbola to one focal distance after the hyperbola.

And, well, just for references, these are the two questions of the hyperbola. I was a little bit careful here to define the axis, and the displacement, and so on, so that the question works. I don't want to belabor that. I let you go back and convince yourself that it is correct, or convince yourself that is not correct, in which case, please let me know so that I can correct it.

But I think I got it right. Anyway, so these are the two equations of the hyperboloids. And this is what in optics is very often referred to as an asphere. Now, this can be slightly confusing.

So I'm going to say it a few times here, and then I'm going to ask you to go back home, read it a few times, then print it out. Post it in front of your bathroom mirror on top of your bed and so on, whatever. So you can see it a few times, and memorize it.

OK, so this type of element whose surface is optimized to give a desired focus and behavior is called in asphere. Again, I have a benefit being Greek. This A in front of the sphere means not a sphere. Give an example from everyday life. I can only think technical terms, and isotropic, aplanatic. Give me an example of from everyday life.

AUDIENCE: Atypical?

GEORGE Atypical, there you are, yeah. Thank you. OK, so it is atypical as in asphere. So another term that is very

BARBASTATHIS: commonly used is in aspheric lens.

Again, I want to emphasize that this kind of a sphere. It works perfectly on axis. But as you can imagine, if you were to move this point object off axis, it would go off. It could create an image, but it would be highly aberrated. It would be subject to, well, definitely spherical and comma possibly. It's a complicated story.

You can imagine that this doesn't happen at the way. There's probably a range over here for point sources at that near enough the axis that this refractor would work actually quite well. So the question is can I make such an element? To answer this question, I should tell you a little bit about how optical elements are usually made.

So what people at least used to do in the old days of expensive optical elements, is they would take a piece of glass. They would mount it on a rotary stage. Then they would bring a diamond tip near the glass. And they would start rotating the glass. And they would rotate basically not at all at the center, and they would let the diamond touch the glass longer and longer as they went away from the tip of the glass. The result of that is, of course, that you create a spherical surface if you move the diamond at a fixed velocity.

So a fixed velocity was very easy. Because all you needed is a model that moves a diamond tip linearly. That's pretty easy to do. To make a hyperbola, you have to plant the path of the diamond tip in a way that results in this nonuniform curvature. So this used to be thought as a very difficult task. Come on in. Someone's outside. Come on in. Don't be shy. She changed her mind.

But in the last 20 years or so, the field of mechatronics that many of you are familiar with then had advanced a lot. So people who are able to design diamond tools that could very carefully plan the path of the diamond tip, and actually create aspheric elements, such as this one. But it still remains very expensive. So it could be an order of magnitude difference in price. If you need to compare a spherical element, compare it to an asphere. Another

Major invention that helped this business is injection molding. Instead of actually making the optical element itself, what people do nowadays is they can make a very expensive negative. And then they inject plastic into the negative. They solidify the plastic, and then they take it out. And they can get any arbitrary surface they want. That's an OK way to make a spheric optical element.

Nevertheless, spherical elements remain very popular. And there's a number of reasons for these, [INAUDIBLE] spheres. They can work very well on axis, but they can have other problems that are perhaps much more severe when they go off axis, which is not a-- it is not a shoe in, so to speak, to use aspheric optical elements. So for that reason, we'll spend quite a bit of time, and also, I guess, because of tradition to analyze what happens to light as it goes through spherical surfaces.

So spherical surface, of course. If you replace the hyperbolas here with a sphere. Let's say I take a sphere. I need to say this very clearly, because the terms can become confusing. If I replace the asphere with a perfect sphere, that is, with a ball.

Then, of course, as you can imagine, what will happen is the wavefront, even inside the sphere will not be planar anymore. You can convince yourself if you think about it that because this sphere actually bends faster than the hyperbola, the arrays that are away from the axes. They will bend more.

So you get a slightly focusing in effect inside the sphere. Because, of course, not a perfect focus. It is itself very high aberrated. And by the time they come out, they become aberrated even more. So you get a very poor quality focus in general. There is a special case where this might not happen, but I will skip that for now. So in general, you get an aberrated image.

So this is what they meant before by aberration. You see that the rays, they form a blur here. They fail to meet at the perfect focus, as it would have been required by the perfect optical system. And this particular type of aberration that occurs when you use a spherical surface or axis in this system is called spherical aberration. Someone asked earlier about spherical. This is spherical.

AUDIENCE: I've got a question.

GEORGE Yes.

BARBASTATHIS:

AUDIENCE: How do you pick z_2 ? Like where do you choose to pick z_2 ? Is it at the focus of the closest rays, or the furthest rays?

GEORGE This came from the hyperbola. z_1 , z_2 , they come from the hyperbola. Depending on the curvature over here,

BARBASTATHIS: define z_1 and z_2 . And then I selected the sphere to match the curvature of the hyperbola near to the axis. So if you look at the animation again. Near the axis, you will see that they actually perfectly match. But then, of course, the sphere carves faster than the hyperbola.

AUDIENCE: I think the question is in the spherically aberrated case where you can not find all the points meeting, how can you define the z_2 ?

GEORGE OK, that's a really good question. So the way you defined z_2 is you look at the two rays that are really, really

BARBASTATHIS: close to the axis. For those two rays that are propagating at the very small angle, then the same z_2 applies.

Then, of course, the rest of the rays. They go off very far from that point. But anyways, z_2 here is exactly the same as it was before. Did that answer your question?

AUDIENCE: Yes. Long

GEORGE Now, if you go off axis, then other types of aberrations kick in. So we're not mentioning them now. We'll come

BARBASTATHIS: back to this topic later, and we find them in more detail. But for now, keep this in mind, that if you use a sphere on axis. Let me repeat. If you use a sphere to focus a point source on axis into a point image, also an axis at the finite distance, then you get spherical aberration in general.

Well, that is true, but if you put it over here. Just move the axis, yes. Well, yeah. OK. Yeah, thank you. Let me define what I mean by the axis here. So typically, film and camera is the elements that we use in order to capture images flat. So if I put here either my camera lens. I should say the tip of the camera, the camera dial, or a film. That defines an axis that is perpendicular to the film.

So Colin is correct, that if I take a point here, and then move it sort of radially through the sphere. It will form a perfect image over here. The fact that I try to detect the image not at this point, but on the plane, means that I will be subject to aberrations. This is the proper definition of aberration. And, of course, even the on axis point will have spherical aberration, as I mentioned earlier. Did I answer that question?

OK, so let's see how we deal with the refraction from a sphere. So [INAUDIBLE], as you can imagine, is a quite complicated problem because here is a ray that is arriving from some point source on axis. It hits the sphere. And then my job is to find which direction this ray will be refracted inside the sphere. So now, again, I can use one of the tools that we have in our arsenal. I can use Fermat's principle, or I can use Snell's law. The two are, of course, identical.

The fact of the matter is in this case, it would be quite complicated to do. So we'll do simplification. First of all, let's define some notation here. I will denote as x the distance of the ray intersects the sphere. I will define as α left the angle that the ray makes with the optical axis to the left of the interface. This is the interface. And then α right is the angle that the ray makes with respect to the horizontal, the optical axis the right hand side of the interface.

OK, that's my notation. And now for generality, I do not assume that this is air anymore, I just assume that I have two dielectrics. One has index n sub left, and the other has index n sub right. It's a little bit cumbersome to carry left and right around, but it is convenient because we will leave that for a mnemonic later on. So how do we solve this problem?

Well, the way I elected to solve it is a little bit easier, it turns out. So for the first time, today I will not use Fermat's principle. I will use Snell's law. And, of course, Snell's law follows from Fermat's principle, so I haven't really changed anything. It's just more convenient to do the math this way. So to apply Snell's law, I have to figure out which way is the perpendicular to the interface at the cross section of the ray with a spherical surface. And, of course, if I have a sphere conveniently, the normal to the sphere goes through the center.

So here is the normal to the sphere. It goes to the center, and some additional rotation. I will denote as R , the radius of the sphere. And ϕ , I will denote the angle that this particular normal, this particular radius makes with, again, with a horizontal axis.

I think this is all the notation that I needed. Actually, I'm not done yet. So I also to apply Snell's law. So to apply Snell's law, I have to look at the angles that enter Snell's law.

So if you recall, we apply Snell relative to the normal to the interface. So here's the normal, here is the ray. Therefore, this angle is θ that will go on the left-hand side of Snell's law. And the corresponding is θ sub right that will go to the right-hand side of Snell's law.

So speaking of which, here is Snell's law-- n left times the sine of the angle on the left-hand side will equal n to the right times the sine of the angle on the right-hand side. And now I have to do a little bit of geometry. So these angles, they are not known to me yet. But I can observe that α left is the same as the angle from the horizontal to the ray over here. And the angle from the horizontal to the radius equals ϕ .

So this means that I have this property over here-- $\theta_{\text{left}} = \alpha_{\text{left}} + 5$ in the corresponding equation for the right-hand side. And I can also substitute these two relationships into Snell's law. I can apply a little bit of trigonometry in order to break the sign of the sum into this rather complicated-looking expression over here.

So I haven't made much progress. So all I managed to do is write a complicated set of trigonometric relationships. In fact, if I were to do a numerical solution to this problem, I would just stop here and feed it into a computer and solve. But I don't want to do that quite yet, because there's quite a bit of intuition to be derived before we go blindly into a computer and plug the numerics.

So of course, the thing that we do in physics when we try to derive intuition from a very complicated formula like this one is we make approximations. And the reasonable approximation to make here is that the rays that are arriving to the element are limited in a relatively small angular range. This turns out to be a very practical assumption, because it is a very common situation in optics to have this small angle approximation.

And it turns out that even when the small angle approximation is not exactly right, you can still do a first pass of the design of an optical system with this approximation. And then you can improve it by doing sort of more complicated numerical analysis. So it's a very good place to start the analysis of an optical system.

So the key assumption here is that α_{left} , this one, is very small. If that is true-- and also, I need one more assumption, which is that this distance over here, the distance that a ray meets the spherical interface, is much smaller than the radius of the sphere. In other words, the sphere is not very small, but it is a relatively small curvature. All right? It has a relatively large radius of curvature or, equivalently, a small curvature.

So the first answers are correct. It is also reasonable to assume that α_{right} , so the right-hand side of interface, is also very small angle. And if I make the set of assumptions, the entire set is called the paraxial approximation. So "paraxial" really means that everything is confined to being near the optical axis.

Very well. So if we do this approximation now, a lot of this gets simplified because of this Taylor series. The sine of the angle of a small angles equals the angle itself. The cosine of a small angle equals 1.

And also, the sine of this angle θ is equal x/r . I don't have to touch that one. This one is actually exact. But I can also neglect the cosine of θ , because-- well, because, again, we are making the paraxial approximation.

So if I do that, and I do a little bit of rearrangement here, I arrive at this explanation that says that the index of the right, the index to the right times the angle to the right equals the index to the left times the angle to the left times this expression over here. You might wonder a number of things. For example, why did I leave the index here? I'm looking for α_{right} . So why don't I just divide along and get done with it?

There is a reason. And it will become apparent in the moment. But for now, I'm almost done, right? I have managed to do what I wanted. I have a pretty closed-form expression for the angle of refraction inside the sphere.

Now I'm going to do something really strange. I will derive a similar expression for light propagating in uniform space. Now, this sounds totally useless, right? But let me do it anyway. And bear with me for a while.

I have a ray which is propagating in free space. For reasons of notation, I will define two refractive indices that are equal. One is n left, one is n right. But they're both equal to the same index of refraction n .

And what I'm interested to do is I'm interested in relating the angle of the ray on the left-hand side with respect to this reference plane over here and the elevation of the ray with respect to the same plane with the same quantities at the second reference plane-- the elevation and the angle of the right-hand side. [? There ?] are other trivial problem, because, for example, Fermat says that in free space, the ray must propagate at a straight line.

So therefore, I can immediately derive that α left equals α right. Not very surprising, right? I'm stating the obvious here, and I'm making a big deal out of it. But nevertheless, let me write it like this. Since the indices are the same, I can also multiply by those indices.

And what about the elevations? Well, the elevations satisfy a simple equation relating the tangent of the angle of propagation multiplied by the propagation distance d . So the elevation to the right equals the elevation to the left times this-- time this quantity.

And then in the spirit of the paraxial approximation again, I'm going to drop the tangent, because the tangent of a small angle equals the angle itself. And I arrive at a set of equations, x right equals x left plus d times α left. And this equation that I got before.

So now let me put the two together. Let me put the free space propagation that I just derived, and relationship between the ray elevation and the ray angle to the left and to the right of a uniform space of distance d , and the equation that I just derived earlier for the refraction of a ray from a spherical surface, spherical refractive surface, of radius r . OK.

So these are the two sets of equations that we got. This is the set of equations that I got immediately before. And this is the set of equations that I saw that I got earlier when I solved the refraction problem at the spherical interface.

And again, I want to emphasize that both of these are correct in the paraxial approximation. These are not exact expressions. The equal sign is a bit of an ambiguous here, because I've sort of assumed that the paraxial approximation is correct. So we're going on with this assumption now. That's why I'm putting equal signs.

OK. And I added one more equation that I didn't write before that says that x to the left equals x to the right. In this case-- and this is a trivial thing-- it says that the ray is continuous. It arrives at the interface at the center point and leaves from the same point. And that's kind of obvious. The ray should not jump.

So what is remarkable about these equations is that they're both fully there, and the relations-- and the quantities that they're linear with respect to are these two elements, the elevation and the angle of the ray. So you can think of them as kind of transport equations. They allow you to transport a ray from two very basic optical elements.

One is free space propagation. It allows you to transport a ray through uniform space. And the other is a refractive interface. Again, it allows you to transport the ray from the left to the right of a refractive interface.

And because they're linear, we can also write them in matrix form as follows. This is not new. I just took these equations and rewrote them. And I rewrote them in a slightly funny way. I kept the index of refraction times the angle.

Can anybody-- now that we see it in this form-- I will go to the matrix in a moment. But now that you see it in this form, can anybody guess why did I go through all this trouble to keep the index of refraction in this expression? Why don't I just write the angle by itself?

Is there a case, a situation that we've seen, where this quantity may be conserved, for example?

AUDIENCE: You have the dispersion thing again. Or where it depends-- where you have different n's, I suppose, and therefore different angles depending on your wavelengths.

GEORGE Yeah. Even simpler, actually. If you are think of Snell's law-- yeah, you're on the right track there. If you think of
BARBASTATHIS: Snell's law, Snell's law says that $n \text{ left } \sin \alpha \text{ left } = n \text{ right } \sin \alpha \text{ right}$. And of course, in the paraxial approximation I drop the sinusoids. So therefore, I arrive to the conclusion that this quantity at the first element of this vector is actually conserved.

And [INAUDIBLE] of course, you can go over it yourselves and convince yourselves that the equation is correct. This is basically just taking the same equation. And these matrices, we'll use them in the next lecture in order to do ray tracing. So basically, an optical element can always be decomposed in a sequence of refractive interfaces and free space or uniform space propagation.

So the reason the matrices are convenient-- sort of jumping ahead a little bit-- is because if you have a cascade of optical elements, then you can ray trace simply by multiplying these matrices. I will do that in detail in the next lecture next Monday. But this is why we went through all this trouble to derive the paraxial approximation and then write it out in this matrix form, so that we can use matrix properties to do ray tracing.

Then there's one more thing before we quit. Do you see Snell's law in any of these laws over here? You can see Snell's law on the handwritten slide over here. But how about-- what about here? Can you see Snell's law somewhere?

Another way to ask my question-- these are questions that I wrote here. Can they capture Snell's law in some way? Do they include it? How?

AUDIENCE: The elevation doesn't change if we apply it at a point the interfaces.

GEORGE OK. That is correct. Yeah, that's correct. OK. Maybe I should restate-- I asked the question in a confusing way.

BARBASTATHIS: The Snell's law at the flat interface-- the same that I rewrote over here-- can I get it somehow from this equation?

What is the radius of curvature of this surface? Infinity, right? So if I plug in infinity, if I plugged r equals infinity, in this case of course, physically what happens is the interface becomes flat. Is the equation over here, this term goes away. And I get what? I get-- $n \text{ left } \alpha \text{ left}, n \text{ right } \alpha \text{ right}, x \text{ left}, x \text{ right}$.

So you write $x \text{ left } = x \text{ right}$, but also $n \text{ left } \alpha \text{ left } = n \text{ right } \alpha \text{ right}$, which is actually Snell's law. So this-- again, I'm restating something that I said before-- this is the reason why we put this product in the top element of the vector when we write this expression.

OK. So we've run out of time. So any questions? Yeah.

AUDIENCE: I think [INAUDIBLE]

GEORGE Oh, you're right. I'm sorry. This would be 1, 0. Yeah. Yeah. I'll fix the notes. There's a typo. The first matrix should
BARBASTATHIS: be 1, 0, and then the rest is correct. Yeah, thank you.

Did everybody hear that? There's a typo in the notes. The elements of the first row of this matrix need to be swapped. OK. So I'll see you all next Monday.