The paraxial approximation in optical system analysis

Douglas A. Kerr

Issue 1
April 5, 2024

ABSTRACT

Rigorous theoretical analysis of an optical system (even a simple lens) involves "ray tracing", in which the path of illustrative "rays" of the light coming to the lens from the "object" being imaged is followed through the various interfaces (*e.g.*, between air and glass) the ray encounters. This involves a number of trigonometric relationships, and thus (especially before the advent of digital computers) was quite tedious to perform.

Then, when the results are in, they often do not paint a "tidy" picture of the system's behavior. For example, for a lens with spherical surfaces, the *focal length* of the lens, a parameter we are often keen to know, is generally not the same if we consider rays passing through the lens at varying distances from its center.

The painful trigonometry can be avoided, and a "tidier" (if somewhat naïve) result obtained, if we limit our consideration to rays that only depart by an infinitesimal angle from lying along the optical axis of the lens (the so-called "paraxial" rays). This article tells the story of doing that. Appendixes give some extended algebraic exercises.

1 THE CONCEPT

Rigorous analysis of the behavior of even a simple lens usually requires formal *ray tracing*, a matter of taking a number of illustrative rays of light and following them across all the interfaces they encounter, using various trigonometric maneuvers and Snell's law of refraction to reckon what happens in each case.

Done wholly by hand, as it had to be when this discipline was first becoming formalized, was extraordinarily tedious. And the results, while rigorous, are occasionally irritating. For example, if we seek to find the focal length of a simple lens with spherical surfaces, we learn that it differs depending on how far from the center of the lens are the rays we choose to be the players in our ray tracing game. (This is not from a flaw in the analysis process, but rather is an inconvenient reality of a lens with spherical surfaces.)

But if we limit ourselves to a cohort of rays that lie infinitesimally close to the optical axis, and whose paths are only at an infinitesimal angle to the optical axis, then the "limit" results can be used to produce a far simpler mathematical model of the process of refraction.

This will not be accurate for any real rays not "almost" lying along the optical axis. But this "fiction" can lead to a useful understanding of how refraction will give a lens its optical properties.

Since this fiction does not give exactly the true picture of the path of any such real rays, it is often spoken of as the "paraxial **approximation**".

2 BACKGROUND

2.1 Introduction

The presentations here revolve around several fundamental concepts of mathematics and optics. For the benefit of the reader who may not be fully familiar with these, I will review them a bit before I proceed

2.2 Limits

The topic of this article in a sense revolves around the mathematical concept of a *limit*.

Suppose we have this equation:

$$y = \frac{(x+1)^2 - 1}{x} \qquad (1)$$

The value of y is of course undefined for $x=0$, since that would put into play the forbidden "division by zero".

But if we let x be some very small value, there is a corresponding value of y, and it is not far from 2. It in fact exactly $(2+x)$. So the closer x is to 0 (but not actually 0), the closer will y be to 2.

We describe this situation thus: "In the limit, as x approaches zero, y approaches 2".

We have a similar thing in trigonometry. We remember that, for an angle $\alpha=0$, $\sin\alpha = 0$ and $\tan\alpha=0$. And for very small values of α, $\sin\alpha$ is still very near α (in radians). The same for $\tan\alpha$.

So we say, "In the limit, as α approaches 0, $\sin\alpha$ approaches α (in radians)". And the same for $\tan\alpha$.

So, if we imagine that we are working in a regime of infinitesimal angles, we can replace $\sin\alpha$ (or $\tan\alpha$) by α (in radians).

A corollary is that, in a regime of infinitesimal angles, we can replace the angle with the horizontal of a line (in radians) by its slope, **u**.

We will draw heavily on this as part of the story here.

2.3 The spherical lens

The classical lens used in optical instruments (initially telescopes, long before there were cameras) has surfaces that are portions of spheres.

It was not that this was the ideal shape for the surface from an optical standpoint, but rather there were straightforward (if tedious) techniques for generating a spherical surface on a glass "blank" that did not depend on having some cutting or grinding tool which itself had somehow been given a nearly-perfect spherical profile. Rather, the tool and the nascent lens each inherently wore themselves to a theoretically-perfect spherical profile.

Not surprisingly, then, the various common means of mathematically predicting the theoretical properties of a lens usually applied to a lens with spherical surfaces.

2.4 Curvature

Important in this area is the concept of the *curvature* of a lens' surface.

In Figure 1 we see a tiny part of a a section through a lens' surface. (In this figure, curvature and angles have been exaggerated for clarity.)

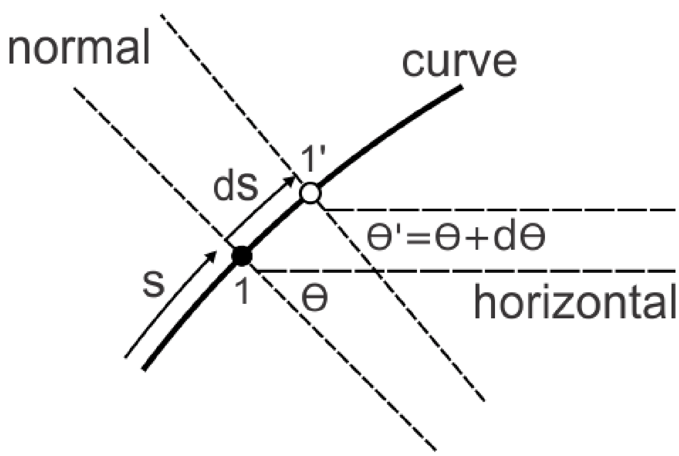


Figure 1.

The *normal* to the curve at any point is a line perpendicular to the tangent to the curve at that point.

I will use the variable **s** to describe a location on the curve. It is measured **along the curve**, from some defined origin.

We consider point 1. The angle its normal makes with our reference direction (here, "horizontal") is Θ. (It is negative in this example.)

Now we move an infinitesimal distance, **ds**, along the curve, to point 1′. The angle of the normal there (Θ′) is changed from Θ by the infinitesimal amount dΘ.

The ratio of the change in the angle of the normal (dΘ) to the distance traveled along the curve to cause it (ds) is spoken of as the *curvature* (C) of the curve at point 1:

$$\frac{d\Theta}{ds} = C \tag{2}$$

Note that in this case dΘ/ds is negative, since as **s** increases Θ decreases. This implies that in this case **C** is negative. Hold that thought.

Note that this definition makes no assumption about which direction the curve is heading at the point of interest (that is, its slope is not involved in the definition).

It can be shown that:

$$C = \frac{1}{R} \tag{3}$$

where **R** is the radius of curvature of the curve at the point of interest. So based on what we saw earlier, in this case it would seem that **R** is negative, which at first seems paradoxical.

But under the convention widely used in formal mathematical work, the radius is essentially defined as the distance from the center of the curve to the curve, and for the part of the curve we are interested in here, the center is a long way to the right, so that distance is to the left, and is thus negative. Thus, in this case, **R** is in fact negative.

2.5 Refraction basics

2.5.1 *Introduction*

In this context, *refraction* refers to the change in the direction of a ray of light when it crosses the boundary between two different transparent materials, as for example when crossing from air into glass at the front surface of a glass lens.

2.5.2 *The index of refraction*

For any transparent medium, an important parameter is its *index of refraction*, which will influence how a ray of light will be diverted in its path when crossing a boundary between two different materials.

In most optical work, the index of refraction is normalized to the index of refraction of empty space (“a vacuum”), which in that scheme is thus defined to have an index of refraction of exactly 1. The index of refraction of air is only very slightly different from that of empty space, and so in most "blackboard" optical work we consider the (normalized) index of refraction of air to be exactly 1.

Just to put things in perspective, the (normalized) index of refraction of most of the different types of glass, and the different types of transparent plastic, used in lenses runs in the range of about 1.50 to 1.80. Often in "blackboard" exercises where some hypothetical lens is being considered, never mind what material it might be, we arbitrarily use an index of refraction of 1.5 to make the calculations handy.

2.5.3 *Snell's law*

The amount by which the direction of a ray of light is changed when passing an interface between two materials of differing index of refraction is given by Snell's[1] law. I will work from this figure:

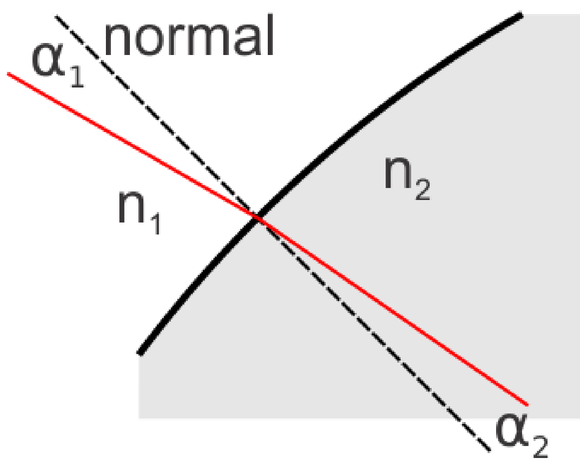


Figure 2.

This is 2-dimensional section of the air-glass interface around a point of interest.

The heavy line represents the boundary between two regions of different transparent materials, with indexes of refraction $\mathbf{n_1}$ and $\mathbf{n_2}$, where $\mathbf{n_2}$ is the greater. I show the second material shaded for identification.

Because in most of our work we will be dealing with curved interfaces (the curved surfaces of lenses), for the sake of realism I have shown a

---

[1] Named in honor of Dutch astronomer Willebrord Snellius (1580–1626), "Snell" for short, who was an important figure in calling attention to this relationship, which however he did not first discover.

curved interface in the figure. But note that the value of the curvature of the interface at the point of interest is of no concern to us. It does not in any way affect the process of refraction of this particular ray.

The dashed black line represents the normal to the interface surface at the point where the ray (itself red) "strikes".

We assume that the arriving ray of interest does not arrive along the normal but rather at angle, α_1, to it. We then note that, having passed across the interface, the ray follows a new direction, this at an angle of α_2 to the normal.

These two angles are related precisely by Snell's law, thus:

$$\frac{\alpha_2}{\alpha_1} = \frac{n_1}{n_2} \tag{4}$$

We can see that there probably will be some trigonometry in our future.

3 THE PARAXIAL APPROXIMATION–WHY IT IS ADVANTAGEOUS

3.1 The paraxial cohort of rays

Suppose we decide to do ray tracing but (through some flash of insight) we decide to consider only rays that lie infinitesimally close to the optical axis, and whose paths are only at an infinitesimal angle to the optical axis. These are spoken of as the paraxial rays; the prefix "par" (we see it in other words as "para") means "alongside".

For conciseness in the discussions, I will refer to this realm of infinitesimals as "Lilliput".

Once we do that, we can justifiably proceed as follows

1. Because we can treat the sine and tangent of an angle as the same as the angle (in radians), we can replace the angle of a line (with respect to our reference direction, here the optical axis) with the slope of the line (considering the optical axis as "horizontal"). I sometimes will refer to such a slopes as the "slope proxy" for an angle.

2. We can replace the angle between two lines with the difference in their slopes.

3.2 A further advantage

There is an additional advantage, which I will discuss using this figure:

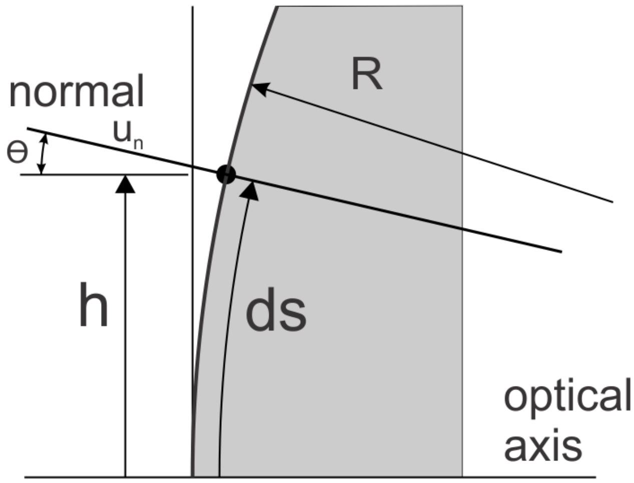


Figure 3.

We saw earlier that the angle of the normal to a curve (Θ) varies, as we travel some perhaps infinitesimal distance along the curve (ds), proportionally to the distance traveled, the constant of that proportionality being defined as the curvature, C, of the curve (which is the reciprocal of the radius, R, of the curve) at the point of interest.

Thus, in Figure 3, to determine Θ we will need to consider the distance **ds**. But, in Lilliput, this is considered to be identical to **h** (the "height" of the point of interest).

Then, since we are able to replace that angle, Θ, with the slope of the normal (u_n), we find that:

$$u_n = \frac{h}{R} \qquad (5)$$

Very handy!

4 THE MECHANICS

An extensive demonstration of how the overall mechanics of ray tracing might proceed in Lilliput is given in Appendix A and Appendix B.

5 SCALING IT UP

Of course, we get limited insight into the behavior of a lens by actually considering the paraxial rays—the unicorns of optics. What is often done is to "scale up" the paths that the paraxial rays would take into paths lying at significant distances from, and angles with, the optical axis. This is mathematically almost trivial to do.

Of course that is not what rays at those locations would actually do—that's why we speak of the paraxial **approximation**.

6 FOR AN ENTIRE LENS

Of course, in an entire lens, we have a first surface (from air to glass) and a little later a second surface (glass to air).

But we can reckon the refraction of the ray at this second surface in essentially the same way as for the first surface

I will spare both of us a detailed example of that!

7 A RELATED FICTION

Even when using the "scaled up" paraxial fiction, in a realistic lens we must deal with the travel of the ray from the first lens surface to the second. Among other things, the height of the ray, **h**, at the second surface will not in general be the same as it was at the first surface

In addition, if we try and simplify the picture by treating the refraction of a ray as occurring in a single stroke, at a single point, where that point is in the lens depends on the distance of the ray from the optical axis. This in turn makes more complicated the reckoning of the journey of the ray beyond the lens.

We can dispose of these nuisances by adopting another fiction, the "thin lens" fiction.

In that fiction we say that even through the lens has, perhaps on both sides, a spherical surface, nevertheless the distance from the first surface to the second surface (in the direction along the optical axis), the "local thickness" of the lens, is everywhere zero.

This is of course impossible from a geometric standpoint. Fictions are like that.

8 ABOUT "DENSITY"

We often see it said, as a reminder of the general way in which refraction occurs, that:

> When crossing an interface to a more-dense material, the ray is refracted toward the normal. When crossing an interface to a less-dense material, the ray is refracted away from the normal.

How does the "density" of the materials enter into this?

Density (in the actual scientific meaning) does not enter into this at all. But as scientific wrings about optics began, the commonest situation was when we considered an interface between air and some kind of glass. It was of course the difference in index of refraction of the two materials that was of consequence, that being greater for glass than for air.

But indeed glass is more “dense” (in the common sense, as well as in the scientific sense) than air, and so it became the custom to give the general guide to refraction in the form stated above. That was perhaps considered more easily understood by those with limited scientific background, rather than the accurate statement, which could be:

> When crossing an interface to a material with a greater index of refraction, the ray is refracted toward the normal. When crossing an interface to a material with a lesser index of refraction, the ray is refracted away from normal.

This is of course a qualitative expression of the implications of Snell’s law.

-#-

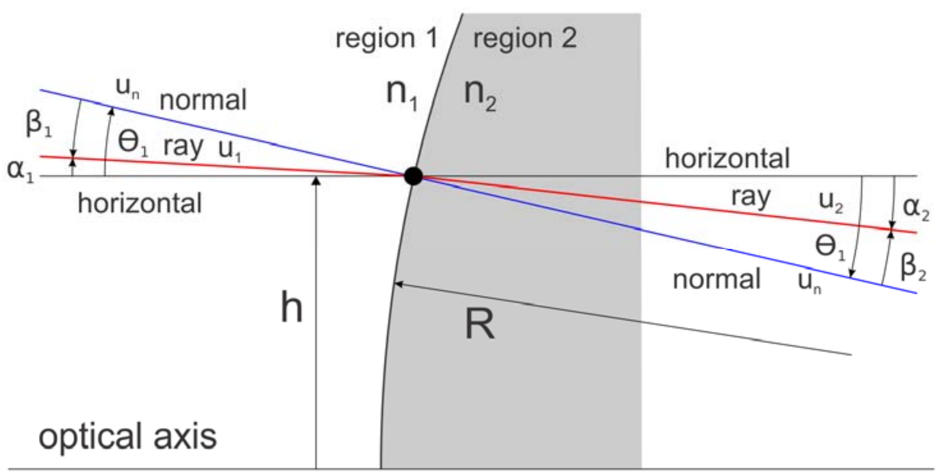
Appendix A
The detailed mechanics

A.1 INTRODUCTION

In this appendix, I follow a ray through one interface, showing how the mathematics will be applied in our (fictional) realm of the paraxial approximation—in our "Lilliput".

A.2 THE LIFE AND TIMES OF A RAY

Consider this figure, which shows the interface between two regions of materials of differing index of refraction (n1 and n2), in a tiny region near the optical axis. This might be at the front face of a spherical lens. In particular, we assume that Region 1 is air, and thus $n_1 = 1$ (we will use this later).



For clarity, curvature, angles, and distances are shown greatly exaggerated. But remember that we are still actually working in the regime of infinitesimally small distances and angles (our Lilliput).

We consider a ray (the red line), with angle to the horizontal α_1 (and slope u_1)[2] hitting the interface at the point marked by the dot. The angle of the normal to the curve at the point of interest (blue) has an angle with the optical axis of Θ_1, and its slope is un. The angle the ray makes with the normal is designated β_1.

As the ray departs from the interface, it makes an angle of β_1 with the normal and an angle of α_1 with the optical axis. And:

$$\beta_1 = \alpha_1 - \Theta_1 \qquad (6)$$

[2] Recall that, in Lilliput, u_1 is considered to be identical to α_1 (expressed in radians)

Snell's law tells us that:

$$\frac{\beta_2}{\beta_1} = \frac{n_1}{n_2} \tag{7}$$

and thus:

$$\beta_2 = \beta_1 \frac{n_1}{n_2} \tag{8}$$

Now:

$$\beta_2 = \alpha_2 - \Theta_1 \tag{9}$$

Because of the interchangeability of angles and slopes in Lilliput:

$$u_{\beta 1} = u_1 - u_n \tag{10}$$

and

$$u_{\beta 2} = u_2 - u_n \tag{11}$$

Thus we can restate Equation 8 as:

$$u_{\beta 2} = u_{\beta 1} \frac{n_1}{n_2} \tag{12}$$

Proceeding, I will assume that we know the radius, R, and thus the curvature, C, of the curve at this place. I will work with C (1/R) as it makes the equations a bit more tidy.

Substituting for $u_{\beta 1}$ and $u_{\beta 2}$, we get

$$u_n - u_2 = (u_1 - u_n) \frac{n_1}{n_2} \tag{13}$$

But we know that, in Lilliput, we can consider:

$$u_n = hC \tag{14}$$

Substituting (and rearranging to make the signs more convenient), we get:

$$u_2 - hC = (hC - u_1) \frac{n_1}{n_2} \tag{15}$$

Solving for u_2 we get:

$$u_2 = \left(hC - u_1\right)\frac{n_1}{n_2} + hC \qquad (16)$$

Then, with $n_1 = 1$, as it is in our example (where the first region is air) we get:

$$u_2 = \frac{(hC - u_1)}{n_2} + hC \qquad (17)$$

no trigonometry having been involved!

-#-

Appendix B
The focal length of a lens

B.1 INTRODUCTION

The focal length s a lens is one of its most fundamental and important parameters. For example, the distance from the lens to an object point, P, and the distance from the lens to the image point that is created from that object point, Q, is given by:

$$\frac{1}{P}+\frac{1}{Q}=\frac{1}{f} \tag{18}$$

where f is the focal length of the lens. (This equation is set up for if we, for simplicity's sake, consider both P and Q to be non-directed distances and thus are both positive.)

In this appendix, I will follow a representative ray though an entire lens, following the laws of Lilliput, and show how that leads us to a view of the focal length of our example lens that is consistent with the familiar rule of thumb[3].

I do this by way of some straightforward but lengthy and somewhat dreadful-looking algebra, to arrive at a wonderfully simple conclusion. The reader not anxious for further algebraic entertainment may wish to jump to the end of this appendix for the *denouement*.

B.2 OUR QUARRY: A HYPOTHETICAL LENS

In an entire lens, we have a first surface (from air to glass) and a little later a second surface (glass to air).

Having analyzed the refraction of our ray at the first surface (a process we saw in Appendix A), we can then reckon its travel to the second surface, where (if u_2 from the first-surface analysis is not 0), the height, h, will be different.

Then we can reckon the refraction of the ray at this second surface in essentially the same way as in the work above.

Here I will do just that. But I cheat in not doing it for the general case. Rather, to make this simpler, I use an especially handy case: a plano-convex lens (having a planar first surface), with a ray from an object point at infinity (which arrives perfectly parallel to the optical axis).

In this figure, we see a tiny part of that plano-convex lens.

---

[3] Yes, a bad pun, given my use of "Lilliput" as a metaphor for the realm of p

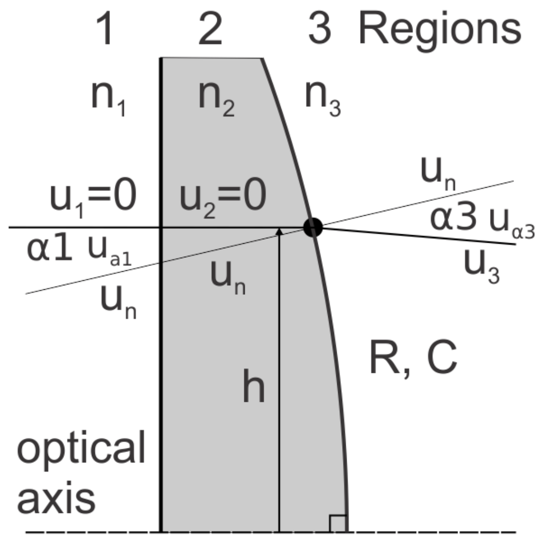


Figure 4. Plano-convex lens

That description means (in this example) a lens whose first surface is a section of a plane and whose second surface is a portion of a convex (outward) sphere (in this example with radius R and thus curvature C). We will see shortly why I chose that lens form for this example.

As before, we will be working in the paraxial regime—in our realm of Lilliput.

Regions 1 and 3 are the air before and after the lens, respectively, and region 2 is the lens proper (glass). Various subscripts follow that notation.

As I said earlier, I will assume an object point “at infinity”, as a result of which the rays from it arrive precisely parallel to the optical axis. On the figure, we see one such ray, and note that its slope (the paraxial proxy for its angle with respect to the optical axis), u_1, is 0.

Since the slope of the normal on the first (planar) surface is everywhere 0, it is 0 where the ray of interest strikes the first surface. Since the slope of the ray and the slope of the normal are the same (that is, the angle between them is 0), the application of Snell’s law tells us that the ongoing ray from this surface also has the same slope as the normal; that is, it has not been deflected by refraction.

In fact, my choice of a lens with a planar first surface for this example is just so that, in a situation that is credible overall, we will have a ray traveling in the glass precisely parallel to the optical axis (with its slope, u_2, being 0).

Now we examine what happens at the second (glass-to-air) interface. At the point where the ray, inside the lens, strikes its second surface (at a distance h from the optical axis), the slope of the normal to that surface at that point, u_n, is given (under the paraxial outlook) by:

$$u_n = hC = \frac{h}{R} \tag{19}$$

This is of course the slope of the normal on both sides of the interface, since it is in fact a line.

Note that, as discussed above in Section 2.4, for this situation **R** is positive, so $\mathbf{u_n}$ will be positive (as seen on the figure).

The angle between the ray, arriving at the second surface, and the normal at that point, $\boldsymbol{\alpha_2}$ (not labeled in the figure), can be represented by its slope proxy, $\mathbf{u_{\alpha 2}}$, given by:

$$u_{\alpha 2} = u_2 - u_n \tag{20}$$

Since, in this example, $\mathbf{u_2}$ is the same as $\mathbf{u_1}$, which is 0, we can write that as:

$$u_{\alpha 2} = -u_n \tag{21}$$

We can replace u_n with H/R, giving:

$$u_{\alpha 2} = -\frac{h}{R} \tag{22}$$

After refraction, the ray exits this surface at an angle with the normal of α_3, which can be represented by its paraxial slope proxy, $u_{\alpha 3}$.

Now by Snell's law (as applies in Lilliput), the slope proxy for the angle between the exiting ray and the normal is given by:

$$u_{\alpha 3} = u_{\alpha 2} \frac{n_2}{n_3} \tag{23}$$

Substituting for $u_{\alpha 2}$, we get:

$$u_{\alpha 3} = -\frac{h}{R} \frac{n_2}{n_3} \tag{24}$$

The slope proxy for the angle of the exiting ray (with respect to the optical axis), u_3, is given by:

$$u_3 = u_n + u_{\alpha 3} \tag{25}$$

So, substituting for $u_{\alpha 3}$, we get:

$$u_3 = u_n - \frac{h}{R}\frac{n_2}{n_3} \tag{26}$$

Then substituting for u_n, we get:

$$u_3 = \frac{h}{R} - \frac{h}{R}\frac{n_2}{n_3} \tag{27}$$

which we can rewrite as:

$$u_3 = \frac{h}{R}\left(1 - \frac{n_2}{n_3}\right) \tag{28}$$

It is now of interest where that existing ray crosses the optical axis (since that is where a point focus of the object point will be formed by this ray and other following the same math).

We see how that works in this figure:

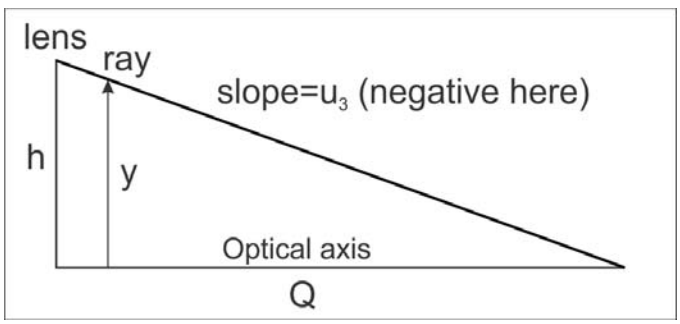


Figure 5.

The altitude of the ray is **y**, and as the ray leaves the lens, **y** = **h**. As we go to the right, the altitude changes at a rate equal to the slope, $\mathbf{u_3}$, which is negative in this case. When it has lost as much altitude as it originally had (**h**), it intersects the optical axis, at a distance Q (the "image distance"). Thus Q is given by:

$$Q = -\frac{h}{u_3} \tag{29}$$

The minus sign is because the altitude, **y**, must **decrease** by the amount **h** for the ray to intersect the optical axis. And, for a negative $\mathbf{u_3}$ (as we have in this example), this results in a positive value of Q, which we can expect based on the figure.

Substituting for u_3 gives us this:

$$Q = -\frac{h}{\frac{h}{R}\left(1 - \frac{n_2}{n_3}\right)} \qquad (30)$$

But in this example n_3 is 1 (for air). So this becomes:

$$Q = -\frac{h}{\frac{h}{R}(1 - n_2)} \qquad (31)$$

which simplifies to:

$$Q = \frac{R}{n_2 - 1} \qquad (32)$$

Hold that thought!

B.3 THE FOCUS EQUATION

For an object point at a distance P in front of the lens, the distance (behind the lens) to the resulting image point is given by:

$$\frac{1}{B} + \frac{1}{Q} = \frac{1}{f} \qquad (33)$$

where **B** is the object distance[4] and **Q** is the image distance.

But with the object distance, **B**, being "infinity" (as for our example), this becomes:

$$\frac{1}{Q} = \frac{1}{f} \qquad (34)$$

which of course is equivalent to

$$Q = f \qquad (35)$$

Substituting in equation 32, we get:

$$f = \frac{R}{n_2 - 1} \qquad (36)$$

[4] Often **P** and **Q** are used for the object and image distances, but we later will use **P** for refractive power, so I arbitrarily use **B** for the object distance here.

B.4 LENS FOCAL LENGTH ANOTHER WAY

An important intermediate result in analyzing the focal length of a lens is the *surface power* of each of its surfaces. This essentially describes the potency of that surface in causing the refraction of a ray crossing it. That value, P_s, is given by:

$$P_s = \frac{n_2 - n_1}{R} \tag{37}$$

where n_1 and n_2 are the indexes of refraction on the first side of the surface and the second side of the surface, respectively.

Then, for an entire lens, with two surfaces, separated by some distance, Gullstrand's equation gives us the refractive power of the entire lens, $\mathbf{P_L}$:

$$P_L = P_1 + P_2 - P_1 P_2 t \tag{38}$$

where $\mathbf{P_1}$ and $\mathbf{P_2}$ are the surface powers of the first and second surfaces, respectively, and **t** is the thickness of the lens.

But in Lilliput we subscribe to the *thin lens fiction* (see Section 7), and so we consider **t** to be 0, and this becomes:

$$P_L = P_1 + P_2 \tag{39}$$

Substituting for the surface powers, that becomes:

$$P_L = \frac{n_2 - n_1}{R_1} + \frac{n_2 - n_1}{R_2} \tag{40}$$

and thus:

$$P_L = (n_2 - n_1)\left(\frac{1}{R_1} + \frac{1}{R_2}\right) \tag{41}$$

But since we know that $n_1 = 1$ (for air), this becomes:

$$P_L = \left(n_2 - 1\right)\left(\frac{1}{R_1} + \frac{1}{R_2}\right) \tag{42}$$

And since in our model R_1 is "infinite" (the first surface being planar), we can simplify that to:

$$P_L = \frac{n_2 - 1}{R_2} \tag{43}$$

The focal length of the lens, f, is by definition given by:

$$f = \frac{1}{P_L} \tag{44}$$

or

$$f = \frac{R}{n_2 - 1} \tag{45}$$

Exactly as given by Equation 36.

Thus we have one demonstration of the credibility of working under the paraxial outlook—working in Lilliput.

-#-