Derivation of Refraction Formulas

Paul S. Heckbert

ABSTRACT

We derive three alternative formulas for the refracted ray direction in ray tracing in order to prove their equivalence and to demonstrate the process of translating physical laws into optimized computational formulas.

It is common knowledge that light rays refract when they strike an interface between two different transparent media, such as air-water, air-glass, or glass-water. In 1621 Dutch mathematician Willebrord Snell discovered a formula quantifying this observation: the ratio of the sines of the incident and refracted angles equals the ratio of the *indices of refraction* of the two materials. Snell's law is:

$$\eta_1 \sin\theta_1 = \eta_2 \sin\theta_2$$

where θ_1 is the angle of incidence, θ_2 is the angle of refraction (both measured from the perpendicular to the interface) and η_1 and η_2 are the two indices of refraction on the incident and refracted sides of the interface, respectively.

Light passing through a material slows relative to its speed in a vacuum by a factor equal to the index of refraction of that material. In fact, Snell's law is a simple consequence of this speed variation and Fermat's *Principle of Least Time*, which states that light takes the fastest path to get from one point to another [Feynman63].

For computation we need to recast Snell's law in terms of (x, y, z) direction vectors. This can be done in several different ways. In the derivations below we make extensive use of angles and trigonometry, but thankfully, it is possible to eliminate all of these terms from the final formulas, so θ_1 and θ_2 need never be computed. As a convention, vectors are upper case and scalars are lower case.

Appeared in Introduction to Ray Tracing, (Andrew Glassner, ed.), Academic Press, London, 1989, pp. 263-293.
This version is missing figures.

Whitted's Method

We first derive the refraction formulas which appeared in Whitted's original paper [Whitted80]. Referring to figure 1, we are given the incident ray direction I and surface normal N, and we need to calculate the transmitted (refracted) ray direction T'. Whitted assumes that N is unit, but not I. First, we scale the incident ray I so that its projection on N is equal to N. Recall from geometry that the component of I parallel to N is $I_{par} = N(I \cdot N)/(N \cdot N)$. But N is a unit vector, so $I_{par} = N(I \cdot N)$. As shown in figure 2, by similar triangles we have

$$\frac{-I}{-I'} = \frac{-I_{par}}{N} = I \cdot N$$

so $I' = I/(-I \cdot N)$. The vector $I' + N$ is thus parallel to the surface (a surface tangent), so we can write the refracted ray as $T' = \alpha(I' + N) - N$ for some α. Note that this refracted ray is not necessarily a unit vector. We must now express α in terms of I, N, and I'. As shown in figure 2, $|I'| = \sec\theta_1$, $|I' + N| = \tan\theta_1$, and $\alpha|I' + N| = \tan\theta_2$, so

$$\alpha = \frac{\tan\theta_2}{\tan\theta_1} = \frac{\sin\theta_2}{\sin\theta_1}\frac{\cos\theta_1}{\cos\theta_2} = \frac{(\eta_1/\eta_2)\cos\theta_1}{\sqrt{1-\sin^2\theta_2}} = \frac{(\eta_1/\eta_2)\cos\theta_1}{\sqrt{1-\eta_1^2/\eta_2^2\sin^2\theta_1}} = \frac{1}{\sqrt{n^2\sec^2\theta_1 - \tan^2\theta_1}}$$

where $n = \eta_2/\eta_1$, employing Snell's law to eliminate the ratio of sines. We can now eliminate the trigonometric terms:

$$\alpha = (n^2|I'|^2 - |I' + N|^2)^{-1/2}$$

Total internal reflection occurs when α is imaginary (square root of a negative number). This happens when rays travel from a dense material to a sparser one ($n < 1$) and the incident angle is above a critical angle: $\theta_1 > \theta_c = \sin^{-1} n$.

In Whitted's article, he used slightly different notation than the above: V for I, V' for I', P for T', k_n for n, and k_f for α.

Heckbert's Method

The second formula comes from [Heckbert-Hanrahan84]. Referring to figure 3, the basic idea is to decompose I into its components parallel and perpendicular to N and then synthesize the transmitted ray T from these components. In this formula we assume that I is unit in addition to N, and we'll guarantee that T will be unit. Since I is unit, $c_1 = \cos\theta_1 = -I \cdot N$. The parallel and perpendicular components of I can then be written: $I_{par} = -c_1 N$ and $I_{perp} = I + c_1 N$. As a bonus, we can easily compute the unit reflected ray direction as shown by the parallelogram in figure 4:

$$R = I + 2c_1 N$$

The refracted ray can be expressed as $T = \sin\theta_2 M - \cos\theta_2 N$ where M is a unit surface tangent vector in the plane of I and N:

$$M \;=\; \frac{I_{perp}}{|I_{perp}|} \;=\; \frac{I + c_1 N}{\sin\theta_1}$$

Therefore,

$$T = \frac{\sin\theta_2}{\sin\theta_1}(I + c_1 N) - \cos\theta_2 N$$

But by Snell's law, the relative index of refraction η is: $\eta = \sin\theta_2 / \sin\theta_1 = \eta_1/\eta_2 = 1/n$, so

$$T = \eta I + (\eta c_1 - c_2) N$$

where $c_2 = \cos\theta_2$. We can easily express $\cos\theta_2$ in terms of known quantities:

$$c_2 \;=\; \cos\theta_2 \;=\; \sqrt{1 - \sin^2\theta_2} \;=\; \sqrt{1 - \eta^2 \sin^2\theta_1} \;=\; \sqrt{1 - \eta^2 (1 - c_1^2)}$$

Total internal reflection occurs when c_2 is imaginary (square root of a negative number).

Other Method

The third formulation is a slight variation on Heckbert's simply replacing η with $1/n$:

$$T \;=\; \eta I + (\eta c_1 - \sqrt{1 - \eta^2 (1 - c_1^2)}) N \;=\; \frac{I}{n} + \frac{c_1 - n\sqrt{1 - (1 - c_1^2)/n^2}}{n} N \;=\; \frac{I + (c_1 - \sqrt{n^2 - 1 + c_1^2}) N}{n}$$

Comparison of Methods

The formulas are all equivalent, of course, so the only advantage of one over another comes from computational speed and perhaps numerical precision. We count the number of arithmetic operations (square roots, divisions, multiplications, and additions/subtractions) required by each method below:

Whitted's Method

$\sqrt{\ }$	/	$\times$	+	
	1			$n = \eta_2/\eta_1$
	3	3	2	$I' = I/(-I \cdot N)$
			3	$J = I' + N$
1	1	8	5	$\alpha = 1/\sqrt{n^2(I' \cdot I') - (J \cdot J)}$
		3	3	$T' = \alpha J - N$
1	3	3	2	$T = T'/\lvert T'\rvert$
2	8	17	15	TOTAL

note: I is not required to be unit in Whitted's method

Heckbert's Method

$\sqrt{\ }$	/	$\times$	+	
	1			$\eta = \eta_1/\eta_2$
		3	2	$c_1 = -I \cdot N$
1		3	2	$c_2 = \sqrt{1 - \eta^2(1 - c_1^2)}$
		7	4	$T = \eta I + (\eta c_1 - c_2)N$
1	1	13	8	TOTAL

note: Heckbert's method uses $\eta = 1/n$, not n

Other Method

$\sqrt{\ }$	/	$\times$	+	
	1			$n = \eta_2/\eta_1$
		3	2	$c_1 = -I \cdot N$
1		2	3	$\beta = c_1 - \sqrt{n^2 - 1 + c_1^2}$
	3	3	3	$T = (I + \beta N)/n$
1	4	8	8	TOTAL

Either the second or third method will be fastest depending on the relative speed of division on a particular machine.

References

Feynman, Richard P., Robert B. Leighton, and Matthew Sands, *The Feynman Lectures on Physics*, Addison-Wesley, Reading, Mass., 1963, p. 26-3.

Heckbert, Paul S., Pat Hanrahan, ''Beam Tracing Polygonal Objects'', *Computer Graphics (SIGGRAPH '84 Proceedings)*, vol. 18, no. 3, July 1984, pp. 119-127.

Whitted, Turner, ''An Improved Illumination Model for Shaded Display'', *CACM*, vol. 23, no. 6, June 1980, pp. 343-349.