Refraction Ray Caluculation without Trig. Function Calls

Ross Beveridge

December 6, 2013 (Updated December 16, 2013)

Here is the derivation of a refraction ray T without any explicit calls to trig. functions. We will do this derivation two ways. First, assuming a surface normal pointing outward and a vector pointing back to the point of origin - the camera on a first call. Second, reversing the direction of the incoming ray so that it is actually the direction of the ray being cast onto this point of intersection with a surface. *Note, the final case below solving for beta was corrected on 12/16/2013.*

Section 1 - Vector pointing back to Camera (or point of origin)

The elements of this computation are as follows. N is the unit length normal to the surface at the point of ray intersection. W the unit length vector pointing back toward the point where the ray being traced originated, the camera pixel if this is the first call in the recursive chain of calls. The indices of refraction μ_i and μ_t are the indices of refraction on the outside and inside of the boundary, where the refraction ray is moving into the new material. The derivation here begins with Snells Law:

> $eq1 := \sin(\theta_i)\,\mu_i = \sin(\theta_t)\,\mu_t : eq1;$

$$\sin(\theta_i)\,\mu_i = \sin(\theta_t)\,\mu_t \tag{1.1}$$

> $tmp := \sin(\theta_t) : \; eq2 := tmp = solve(eq1, tmp) : eq2;$

$$\sin(\theta_t) = \frac{\sin(\theta_i)\,\mu_i}{\mu_t} \tag{1.2}$$

For simplicity, define

> $eq3 := \mu = \frac{\mu_i}{\mu_t} : eq3;$

$$\mu = \frac{\mu_i}{\mu_t} \tag{1.3}$$

> $eq4 := \sin(\theta_i)\cdot\mu = \sin(\theta_t) : eq4;$

$$\sin(\theta_i)\,\mu = \sin(\theta_t) \tag{1.4}$$

Square both sides

> $eq5 := lhs(eq4)\cdot lhs(eq4) = rhs(eq4)\cdot rhs(eq4) : eq5;$

$$\sin(\theta_i)^2\,\mu^2 = \sin(\theta_t)^2 \tag{1.5}$$

Recognize this step introduces 'false' solutions in the sense that we may have to choose between a solution and its signed opposite. Finally, so we can replace cosines with dot products of terms we know, shift to the cosine squared form of the same equation.

> $eq6 := subs\left(\sin(\theta_i)^2 = 1 - \cos(\theta_i)^2, \sin(\theta_t)^2 = 1 - \cos(\theta_t)^2, eq5\right) : eq6;$

(1.6)

$$\left(1-\cos\left(\theta_i\right)^2\right)\mu^2=1-\cos\left(\theta_t\right)^2 \tag{1.6}$$

Now let us define the refraction ray T as a weighted sum of the normal N and the ray W.

> $eq7 := T = a \cdot W + b \cdot N : eq7;$

$$T = a\,W + b\,N \tag{1.7}$$

>

The cosines now become dot products, $\cos\left(\theta_i\right) = W \bullet N$ and $\cos\left(\theta_t\right) = -T \bullet N$. Further, the refraction ray is expressed as a weighted sum above, so $\cos\left(\theta_t\right) = -N \bullet \left(a\,W + b\,N\right)$. Since the dot produce of the normal with itself is 1, this further simplifies to $\cos\left(\theta_t\right) = -aN \bullet \mathrm{W} - \mathrm{b}$.

To simplify what follows, let us name the scalar resulting from the dot product of N and W as $wn = N \bullet W$.

> $eq8 := \cos\left(\theta_i\right) = \mathrm{wn} :\ \mathrm{eq8};$

$$\cos\left(\theta_i\right) = wn \tag{1.8}$$

> $eq9 := \cos\left(\theta_t\right) = -\mathrm{a} \cdot \mathrm{wn} - \mathrm{b} : \mathrm{eq9};$

$$\cos\left(\theta_t\right) = -a\,wn - b \tag{1.9}$$

Now, substituting these cosine squared expressions in terms of wn back into the squared form of Snells Law (eq6) gives rise to

> $eq10 := subs\left(\mathrm{eq8}, \mathrm{eq9}, \mathrm{eq6}\right) : eq10;$

$$\left(1 - wn^2\right)\mu^2 = 1 - \left(-a\,wn - b\right)^2 \tag{1.10}$$

>

This gives us one of the two quadratic equations we need in order to solve for a and b. The other comes from adding the constraint that T be of unit length.

> $eq11 := \left(a\,W + b\,N\right) \cdot \left(a\,W + b\,N\right) = 1 : \mathrm{eq11};$
> $\mathrm{expand}\left(eq11\right);$

$$\left(a\,W + b\,N\right)^2 = 1$$

$$a^2\,W^2 + 2\,a\,W\,b\,N + b^2\,N^2 = 1 \tag{1.11}$$

>

Recall there are actually dot products and scalar multiplies in this result and also that both vectors are unit length, thus we can simplify this equation

> $eq12 := a^2 + 2\,a\;\;b\,wn + b^2 = 1 : eq12;$

$$a^2 + 2\,a\,b\,wn + b^2 = 1 \tag{1.12}$$

To reiterate our two constraints, they are:

> $eq13 := -1 \cdot simplify\left(eq10\right) : eq13;$

$$\left(-1 + wn^2\right)\mu^2 = -1 + a^2\,wn^2 + 2\,a\,b\,wn + b^2 \tag{1.13}$$

> $eq12;$

$$a^2 + 2\,a\,b\,wn + b^2 = 1 \tag{1.14}$$

With a bit of inspection, note that the second constraint may be written as:

> $eq15 := 2\,a\,b\,wn + b^2 = 1 - a^2 : eq15;$

(1.15)

$$2\,a\,b\,wn + b^2 = 1 - a^2 \tag{1.15}$$

\>

The left hand side may now be substituted into the first equation

\> $eq16 := \left(-1 + wn^2\right)\mu^2 = -1 + a^2\,wn^2 + 1 - a^2 : eq16;$

$$\left(-1 + wn^2\right)\mu^2 = a^2\,wn^2 - a^2 \tag{1.16}$$

\>

This may be essentially solved by inspection once the left hand side and right hand side are simplified and rearranged.

\> $eq17 := \left(wn^2 - 1\right)\mu^2 = a^2\left(wn^2 - 1\right) : eq17;$

$$\left(-1 + wn^2\right)\mu^2 = a^2\left(-1 + wn^2\right) \tag{1.17}$$

\>

Thus, there are two possible solutions for a, the weight applied to W.

\> $eq18 := a = solve(eq17, a) : eq18;$

$$a = (\mu, -\mu) \tag{1.18}$$

\>

Choosing between the two alternatives becomes easy when one remembers that W is pointing away from the direction of the refraction ray, therefore the creation of a refraction ray moving into the material will require a flipping of the direction of W. To reinforce this choice, realize that if μ is equal to one an no refraction bending is to take place then T must be the signed opposite of W. This constraint demands we choose $a = -\mu$.

\> $eq19 := a = -\mu : eq19;$

$$a = -\mu \tag{1.19}$$

Now, once we decide upon using the $a = -\mu$ solution, there are then two choices for b which may be found by solving either of our two original quadratic equations. We will choose the second;

\> $eq20 := subs(\mathrm{eq19}, eq12) : eq20;$

$$\mu^2 - 2\,\mu\,b\,wn + b^2 = 1 \tag{1.20}$$

\> $eq21 := b = solve(eq20, b) : eq21;$

$$b = \left(\mu\,wn + \sqrt{\mu^2\,wn^2 - \mu^2 + 1},\ \mu\,wn - \sqrt{\mu^2\,wn^2 - \mu^2 + 1}\right) \tag{1.21}$$

\>

In choosing between these two alternatives a similar argument may be made to that for choosing a. Consider again the case μ is equal to one and no refraction bending is to take place. The first choice above will not go to zero in this case, therefore it is not the choice we want. Conversely, the second choice nicely goes to zero for the μ is equal to one case. So, in conclusion,

\> $eq22 := b = solve(eq20, b)[2] : eq22;$

$$b = \mu\,wn - \sqrt{\mu^2\,wn^2 - \mu^2 + 1} \tag{1.22}$$

\>

As should now be clear, the overall structure of the solution changes little if one were to shift the direction of the ray W. However, the differences are enough to make a huge practical difference, and getting the relative signs correct of course matters. Thus, the above derivation should may be used as a cookbook, but better to understand all the steps. Also, below, for the sake of completeness, we take this derivation and adjust it to the case of W going in the opposite direction.

Section 2 - Ray Intersecting the Surface

Because this second derivation is built by adjusting the first, it is perhaps a bit verbose and overly complete. However, that is preferable to the option of not having this version at all.

The elements of this computation are as follows. N is the unit length normal to the surface at the point of ray intersection. W the unit length vector indicating the direction of the ray being cast. The indices of refraction μ_i and μ_t are the indices of refraction on the outside and inside of the boundary, where the refraction ray is moving into the new material. The derivation here begins with Snells Law:

> $eq1 := \sin(\theta_i)\,\mu_i = \sin(\theta_t)\,\mu_t : eq1;$

$$\sin(\theta_i)\,\mu_i = \sin(\theta_t)\,\mu_t \qquad \textbf{(2.1)}$$

> $tmp := \sin(\theta_t) : \ eq2 := tmp = solve(eq1, tmp) : eq2;$

$$\sin(\theta_t) = \frac{\sin(\theta_i)\,\mu_i}{\mu_t} \qquad \textbf{(2.2)}$$

For simplicity, define

> $eq3 := \mu = \frac{\mu_i}{\mu_t} : eq3;$

$$\mu = \frac{\mu_i}{\mu_t} \qquad \textbf{(2.3)}$$

> $eq4 := \sin(\theta_i)\cdot\mu = \sin(\theta_t) : eq4;$

$$\sin(\theta_i)\,\mu = \sin(\theta_t) \qquad \textbf{(2.4)}$$

Square both sides

> $eq5 := lhs(eq4)\cdot lhs(eq4) = rhs(eq4)\cdot rhs(eq4) : eq5;$

$$\sin(\theta_i)^2\,\mu^2 = \sin(\theta_t)^2 \qquad \textbf{(2.5)}$$

Recognize this step introduces 'false' solution in the sense that we may have to choose between a solution and its signed opposite. Finally, so we can replace cosines with dot products of terms we know, shift to the cosine squared form of the same equation.

> $eq6 := subs\left(\sin(\theta_i)^2 = 1 - \cos(\theta_i)^2, \sin(\theta_t)^2 = 1 - \cos(\theta_t)^2, eq5\right) : eq6;$

$$\left(1-\cos\left(\theta_i\right)^2\right)\mu^2 = 1-\cos\left(\theta_t\right)^2 \quad \textbf{(2.6)}$$

Now let us define the refraction ray T as a weighted sum of the normal N and the ray W.

> $eq7 := T = a\cdot W + b\cdot N : eq7;$

$$T = a\,W + b\,N \quad \textbf{(2.7)}$$

>

The cosines now become dot products, $\cos\left(\theta_i\right) = -W \bullet N$ and $\cos\left(\theta_t\right) = -T \bullet N$. Further, the refraction ray is expressed as a weighted sum above, so $\cos\left(\theta_t\right) = -N \bullet \left(a\,W + b\,N\right)$. Since the dot produce of the normal with itself is 1, this further simplifies to $\cos\left(\theta_t\right) = -aN \bullet \mathrm{W} - \mathrm{b}$.

To simplify what follows, let us name the scalar resulting from the dot product of N and W as $wn = N \bullet W$.

> $eq8 := \cos\left(\theta_i\right) = -\mathrm{wn} : \mathrm{eq8};$

$$\cos\left(\theta_i\right) = -wn \quad \textbf{(2.8)}$$

> $eq9 := \cos\left(\theta_t\right) = -\mathrm{a}\cdot\mathrm{wn} - \mathrm{b} : \mathrm{eq9};$

$$\cos\left(\theta_t\right) = -a\,wn - b \quad \textbf{(2.9)}$$

Now, substituting these cosine square expression in terms of wn back into the squared form of Snells Law (eq6) gives rise to

> $eq10 := subs\left(\mathrm{eq8}, \mathrm{eq9}, \mathrm{eq6}\right) : eq10;$

$$\left(1-wn^2\right)\mu^2 = 1-\left(-a\,wn-b\right)^2 \quad \textbf{(2.10)}$$

>

This gives us one of the two quadratic equations we need in order to solve for a and b. The other comes from adding the constraint that T be of unit length.

> $eq11 := \left(a\,W + b\,N\right)\cdot\left(a\,W + b\,N\right) = 1 : \mathrm{eq11};$
> $\mathrm{expand}\left(eq11\right);$

$$\left(a\,W + b\,N\right)^2 = 1$$

$$a^2\,W^2 + 2\,a\,W\,b\,N + b^2\,N^2 = 1 \quad \textbf{(2.11)}$$

>

Recall there are actually dot products and scalar multiplies in this result and also that both vectors are unit length, thus we can simplify this equation

> $eq12 := a^2 + 2\,a\;\,b\,wn + b^2 = 1 : eq12;$

$$a^2 + 2\,a\,b\,wn + b^2 = 1 \quad \textbf{(2.12)}$$

To reiterate our two constraints, they are:

> $eq13 := -1\cdot simplify\left(eq10\right) : eq13;$

$$\left(-1+wn^2\right)\mu^2 = -1 + a^2\,wn^2 + 2\,a\,b\,wn + b^2 \quad \textbf{(2.13)}$$

> $eq12;$

$$a^2 + 2\,a\,b\,wn + b^2 = 1 \quad \textbf{(2.14)}$$

With a bit of inspection, note that the second constraint may be written as:

> $eq15 := 2\,a\,b\,wn + b^2 = 1 - a^2 : eq15;$

(2.15)

$$2\,a\,b\,wn + b^2 = 1 - a^2 \tag{2.15}$$

>

The left hand side may now be substituted into the first equation

> $eq16 := \left(-1 + wn^2\right)\mu^2 = -1 + a^2\,wn^2 + 1 - a^2 : eq16;$

$$\left(-1 + wn^2\right)\mu^2 = a^2\,wn^2 - a^2 \tag{2.16}$$

>

This may be essentially solved by inspection once the left hand side and right hand side are simplified and rearranged.

> $eq17 := \left(wn^2 - 1\right)\mu^2 = a^2\left(wn^2 - 1\right) : eq17;$

$$\left(-1 + wn^2\right)\mu^2 = a^2\left(-1 + wn^2\right) \tag{2.17}$$

>

Thus, there are two possible solutions for a, the weight applied to W.

> $eq18 := a = solve(eq17, a) : eq18;$

$$a = (\mu, -\mu) \tag{2.18}$$

>

If you have a close eye for detail, you will have noticed that changing only the sign on W made no difference in the pair of quadratic equations because the only place it mattered was when we needed the square of the cosine of the initial angle θ_i.

To head off one possible objection, not we do not want to modify our definition of T , in particular no change to signs when we encounter the dot product of N and W in the definition of T should be made. In fact, it is important we keep it as it appears in the first computaton because the a and b we solve for must be proper for the direction of W we've chosen in this version of the derivation.

However, not comes the key observation. We must choose one of the options for a and the reasoning is now the opposite of the case above. In other words, remember W is now pointing into the material and our refraction ray T should point roughly in the same direction. Hence, we now choose $a = \mu$.

> $eq19 := a = \mu : eq19;$

$$a = \mu \tag{2.19}$$

Once we decide upon using the $a = -\mu$ solution, there are then two choices for b which may be found by solving either of our two original quadratic equations. We will choose the second;

> $eq20 := subs(eq19, eq12) : eq20;$

$$\mu^2 + 2\,\mu\,b\,wn + b^2 = 1 \tag{2.20}$$

> $eq21 := b = solve(eq20, b) : eq21;$

$$b = \left(-\mu\,wn + \sqrt{\mu^2\,wn^2 - \mu^2 + 1},\ -\mu\,wn - \sqrt{\mu^2\,wn^2 - \mu^2 + 1}\right) \tag{2.21}$$

>

Changes start here ...

Let us again consider the case μ is equal to one and no refraction bending is to take place. By the same argument made above, namely that b should go to zero in this case. Recall now that wn is a negative number! Therefore, the second choice is better if we assome that we are taking the positive root of wn

squared. So b we now make the same choice from before. (Note, this should be further justified by a worked example! 12/16/13 Ross). Namely,

```
> eq22 := b = solve(eq20, b)[2] : eq22;
```

$$b = -\mu\, wn - \sqrt{\mu^2\, wn^2 - \mu^2 + 1} \tag{2.22}$$

```
>
```