Ray Tracing

Brian Curless
CSE 457
Spring 2014

Reading

Required:

- Shirley, section 10.1-10.7 (online handout)
- Triangle intersection (online handout)

Further reading:

- Shirley errata on syllabus page, needed if you work from his book instead of the handout, which has already been corrected.
- T. Whitted. An improved illumination model for shaded display. Communications of the ACM 23(6), 343-349, 1980.
- A. Glassner. An Introduction to Ray Tracing. Academic Press, 1989.
- K. Turkowski, "Properties of Surface Normal Transformations," Graphics Gems, 1990, pp. 539-547.

Geometric optics

Modern theories of light treat it as both a wave and a particle.

We will take a combined and somewhat simpler view of light – the view of **geometric optics**.

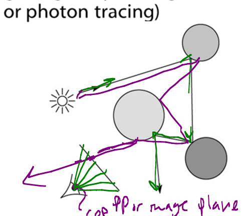
Here are the rules of geometric optics:

- Light is a flow of photons with wavelengths. We'll call these flows "light rays."
- Light rays travel in straight lines in free space.
- Light rays do not interfere with each other as they cross.
- Light rays obey the laws of reflection and refraction.
- Light rays travel from the light sources to the eye, but the physics is invariant under path reversal (reciprocity).

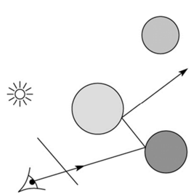
Eye vs. light ray tracing

Where does light begin?

At the light: light ray tracing (a.k.a., forward ray tracing or photon tracing)



At the eye: eye ray tracing (a.k.a., backward ray tracing)

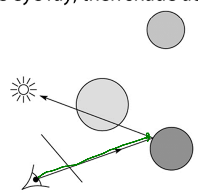


We will generally follow rays from the eye into the scene.

Precursors to ray tracing

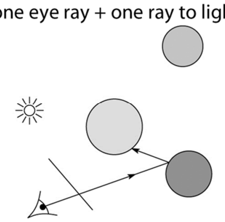
Local illumination

- Cast one eye ray, then shade according to light



Appel (1968)

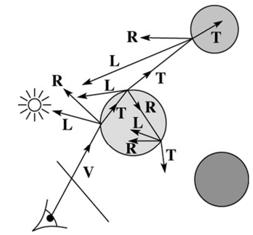
- Cast one eye ray + one ray to light



Whitted ray-tracing algorithm

In 1980, Turner Whitted introduced ray tracing to the graphics community.

- Combines eye ray tracing + rays to light
- Recursively traces rays



Algorithm:

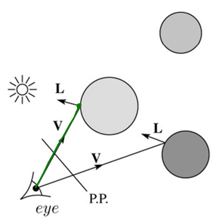
1. For each pixel, trace a **primary ray** in direction **V** to the first visible surface.

2. For each intersection, trace **secondary rays**:

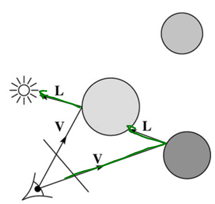
- **Shadow rays** in directions $\mathbf{L_i}$ to light sources
- **Reflected ray** in direction **R**.
- **Refracted ray** or **transmitted ray** in direction **T**.

Whitted algorithm (cont'd)

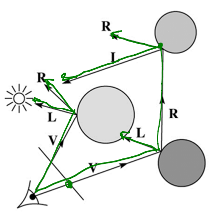
Let's look at this in stages:



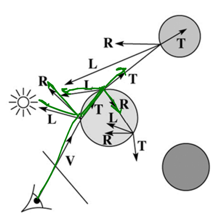
Primary rays



Shadow rays



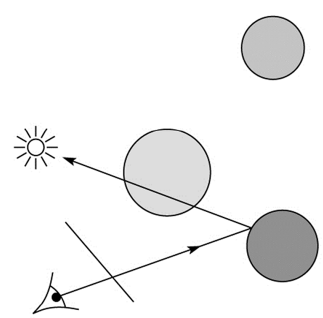
Reflection rays



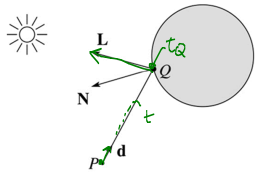
Refracted rays

Ray casting and local illumination

Now let's actually build the ray tracer in stages.
We'll start with ray casting and local illumination:



Direct illumination



A ray is defined by an origin P and a unit direction $\mathbf{d}$ and is parameterized by $t > 0$:

$$\mathbf{r}(t) = P + t\mathbf{d}$$

Let $I(P, \mathbf{d})$ be the intensity seen along a ray. Then:

$$I(P, \mathbf{d}) = I_{\text{direct}}$$

where

- I_{direct} is computed from the Blinn-Phong model

Shading in "Trace"

The Trace project uses a version of the Blinn-Phong shading equation we derived in class, with two modifications:

- Distance attenuation is clamped to be at most 1:

$$A_j^{dist} = \min\left\{1, \frac{1}{a_j + b_j r_j + c_j r_j^2}\right\}$$

- Shadow attenuation A^{shadow} is included.

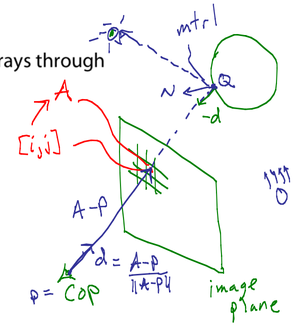
Here's what it should look like:

$$I = k_e + k_a I_{La} + \sum_j A_j^{shadow} A_j^{dist} I_{L,j} B_j \left[k_d \left(\mathbf{N} \cdot \mathbf{L}_j\right) + k_s \left(\mathbf{N} \cdot \mathbf{H}_j\right)^{n_s} \right]$$

This is the shading equation to use in the Trace project!

Ray-tracing pseudocode

We build a ray traced image by casting rays through each of the pixels.



function *traceImage* (scene):

  for each pixel (i,j) in image

    A = *pixelToWorld*(i,j)

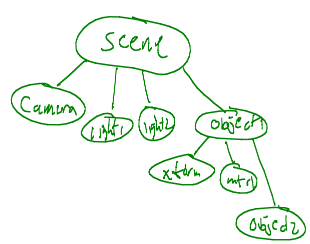
    P = **COP**

    d = (*A* - *P*)/|| *A* – *P*||

    I(i,j) = *traceRay*(scene, *P*, **d**)

  end for

end function



function *traceRay*(scene, *P*, **d**):

  (t, **N**, mtrl) ← scene.*intersect* (*P*, **d**)

  Q ← ray (*P*, **d**) evaluated at t   Q = P + td

  I = *shade*(N, -d, Scene, mtrl, Q)

  return I

end function

Shading pseudocode

Next, we need to calculate the color returned by the *shade* function.

function *shade*(mtrl, scene, *Q*, **N**, **d**):

  I ← mtrl.k_e + mtrl. k_a * I_{La}

  for each light source Light **do**:

    atten = Light -> *distanceAttenuation*(Q)

    L = Light -> getDirection (Q)

    I ← I + atten*(diffuse term + specular term)

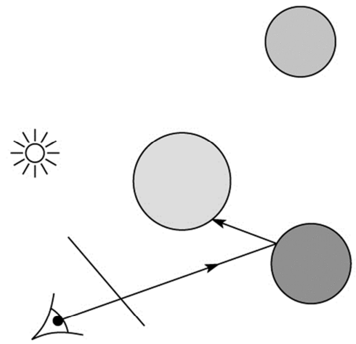
  end for

  return I

end function

Ray casting with shadows

Now we'll add shadows by casting shadow rays:



Shading with shadows

To include shadows, we need to modify the shade function:

function *shade*(mtrl, scene, *Q*, **N**, **d**):

  I ← mtrl.k_e + mtrl. k_a * I_{La}

  for each light source Light **do**:

    atten = Light -> *distanceAttenuation*(*Q*) *

      Light -> *shadowAttenuation*(Q, scene)

    L = Light -> getDirection (Q)

    I ← I + atten*(diffuse term + specular term)

  end for

  return I

end function

Shadow attenuation

Computing a shadow can be as simple as checking to see if a ray makes it to the light source.

For a point light source:

function *PointLight*::*shadowAttenuation*(scene, *P*)

  d = getDirection(P)

  (t, **N**, mtrl) ← scene.*intersect*(*P*, **d**)

  Compute t_{light}

  if (t < t_{light}) **then**:

    atten = (0, 0, 0)

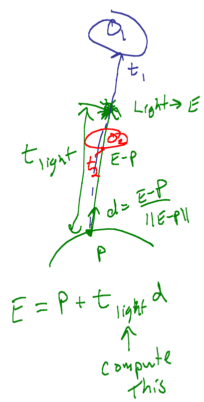
  else

    atten = (1, 1, 1)

  end if

  return atten

end function

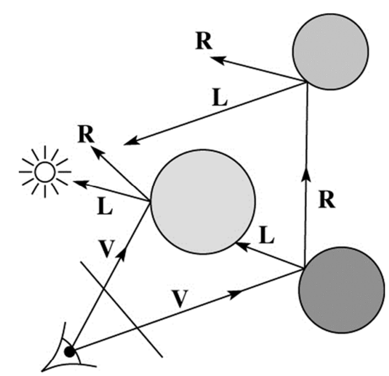


Note: we will later handle color-filtered shadowing, so this function needs to return a *color* value.

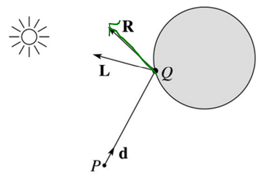
For a directional light, $t_{light} = \infty$.

Recursive ray tracing with reflection

Now we'll add reflection:



Shading with reflection



Let $I(P, \mathbf{d})$ be the intensity seen along a ray. Then:

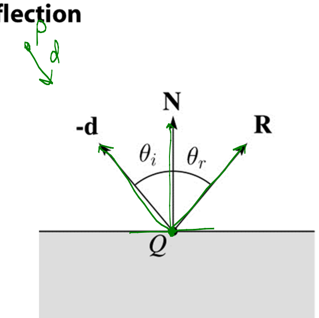
$$I(P, \mathbf{d}) = I_{\text{direct}} + I_{\text{reflected}}$$

where

- I_{direct} is computed from the Blinn-Phong model, plus shadow attenuation
- $I_{\text{reflected}} = k_r I(Q, \mathbf{R})$

Typically, we set $k_r = k_s$. (k_r is a color value.)

Reflection



Law of reflection:

$$\theta_i = \theta_r$$

R is co-planar with **d** and **N**.

Ray-tracing pseudocode, revisited

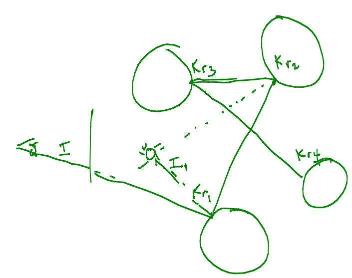
```
function traceRay(scene, P, d):
    (t, N, mtrl) ← scene.intersect (P, d)
    Q ← ray (P, d) evaluated at t
    I = shade(scene, mtrl, Q, N, -d)
    R = reflectDirection( N, -d )
    I ← I + mtrl.kr * traceRay(scene, Q, R)
    return I
end function
```


Terminating recursion

Q: How do you bottom out of recursive ray tracing?

Possibilities:

implement this → max # of levels of recursion



$I = I_1 + k_{r_1}(I_2 + k_{r_2}(I_3 + k_{r_3}(\ldots$

# of bounces

$\prod_{i=0} k_{r_i} < \text{thresh}$

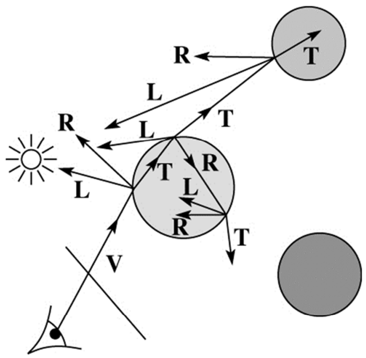
$\Rightarrow$ terminate

early (adaptive) ray termination

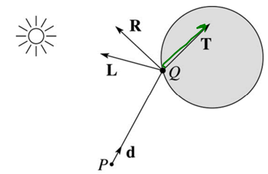
Extra credit

Whitted ray tracing

Finally, we'll add refraction, giving us the Whitted ray tracing model:



Shading with reflection and refraction



Let $I(P, \mathbf{d})$ be the intensity seen along a ray. Then:

$$I(P, \mathbf{d}) = I_{\text{direct}} + I_{\text{reflected}} + I_{\text{transmitted}}$$

where

- I_{direct} is computed from the Blinn-Phong model, plus shadow attenuation
- $I_{\text{reflected}} = k_r I(Q, \mathbf{R})$
- $I_{\text{transmitted}} = k_t I(Q, \mathbf{T})$

Typically, we set $k_r = k_s$ and $k_t = 1 - k_s$ (or (0,0,0), if opaque, where k_t is a color value).

[Generally, k_r and k_t are determined by "Fresnel reflection," which depends on angle of incidence and changes the polarization of the light. This is discussed in Shirley's textbook and can be implemented for extra credit.]

Refraction

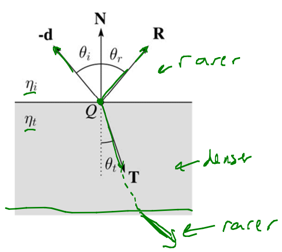
P
d

Snell's law of refraction:

$$\eta_i \sin\theta_i = \eta_t \sin\theta_t$$

where η_i, η_t are **indices of refraction**.

In all cases, **R** and **T** are co-planar with **d** and **N**.

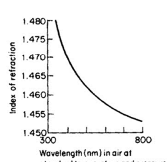


The index of refraction is material dependent.

It can also vary with wavelength, an effect called **dispersion** that explains the colorful light rainbows from prisms. (We will generally assume no dispersion.)

use this outside of objects

Medium	Index of refraction
Vaccum	1
Air	1.0003
Water	1.33
Fused quartz	1.46
Glass, crown	1.52
Glass, dense flint	1.66
Diamond	2.42



Index of refraction variation for fused quartz

Total Internal Reflection

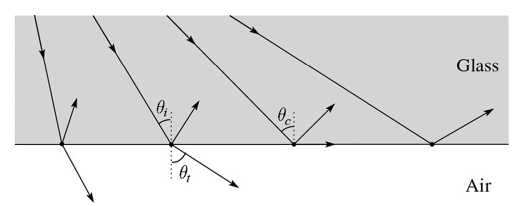
The equation for the angle of refraction can be computed from Snell's law:

$$\sin\theta_t = \frac{\eta_i}{\eta_t}\sin\theta_i \qquad \theta_t = \sin^{-1}\left[\frac{\eta_i}{\eta_t}\sin\theta_i\right]$$

What happens when $\eta_i > \eta_t$?

When θ_t is exactly 90°, we say that θ_i has achieved the "critical angle" θ_c.

For $\theta_i > \theta_c$, *no rays are transmitted*, and only reflection occurs, a phenomenon known as "total internal reflection" or TIR.



Shirley's notation

Shirley uses different symbols. Here is the translation between them:

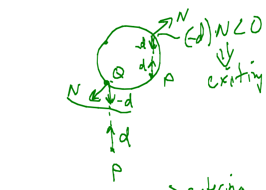
$$\mathbf{r} = \mathbf{R}$$
$$\mathbf{t} = \mathbf{T}$$
$$\phi = \theta_t$$
$$\theta = \theta_r = \theta_i$$
$$n = \eta_i$$
$$n_t = \eta_t$$

Also, Shirley has two important errors that have already been corrected in the handout.

But, if you're consulting the original 2005 text, be sure to refer to the errata posted on the syllabus and on the project page for corrections.

Ray-tracing pseudocode, revisited

```
function traceRay(scene, P, d):
    (t, N, mtrl) ← scene.intersect (P, d)
    Q ← ray (P, d) evaluated at t
    I = shade(scene, mtrl, Q, N, -d)
    R = reflectDirection(N, -d)
    I ← I + mtrl.k_r * traceRay(scene, Q, R)
    if ray is entering object then
        n_i = index_of_air
        n_t = mtrl.index
    else
        n_i = mtrl.index
        n_t = index_of_air
    if (notTIR ( n_i, n_t, N, -d )) then
        T = refractDirection ( n_i, n_t, N, -d )
        I ← I + mtrl.k_t * traceRay(scene, Q, T)
    end if
    return I
end function
```



$(-d)\cdot N \geq 0 \Rightarrow$ entering

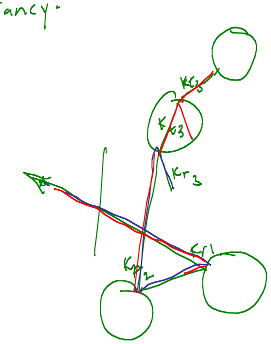
Q: How do we decide if a ray is entering the object?

Terminating recursion, incl. refraction

Q: *Now* how do you bottom out of recursive ray tracing?

Simple: max # of bounces

Fancy:

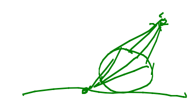


$$\prod_{i=0}^{\#\text{ of bounce}} k_{\{r_i \text{ or } t_i\}} < \text{thresh}$$

$\Rightarrow$ terminate

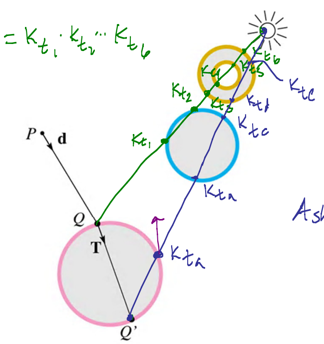
Shadow attenuation (cont'd)

Q: What if there are transparent objects along a path to the light source?



$A_{shad.} = k_{t_1} \cdot k_{t_2} \cdots k_{t_6}$

Implement this

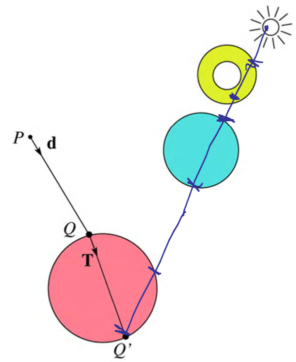


$A_{shad} = k_{t_a} \cdot k_{t_b} \cdots k_{t_e}$

We'll take the view that the color is really only at the surface, like a glass object with a colored transparency coating on it. In this case, we multiply in the transparency constant, k_t, every time an object is entered or exited, possibly more than once for the same object.

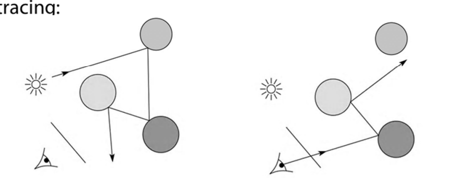
Shadow attenuation (cont'd)

Another model would be to treat the glass as solidly colored in the interior. Shirley's textbook describes a the resulting volumetric attenuation based on Beer's Law, which you can implement for extra credit.

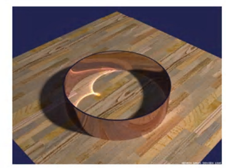


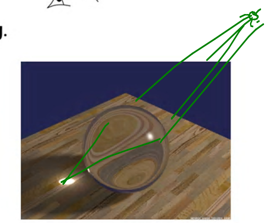
Photon mapping

Combine light ray tracing (photon tracing) and eye ray tracing:



…to get **photon mapping**.





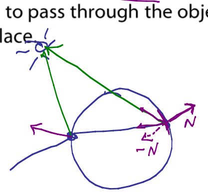
Renderings by Henrik Wann Jensen: http://graphics.ucsd.edu/~henrik/images/caustics.html

Normals and shading when inside

When a ray is inside an object and intersects the object's surface on the way out, the normal will be pointing ***away*** from the ray (i.e., the normal always points to the outside by default).

You must ***negate*** the normal before doing any of the shading, reflection, and refraction that follows.

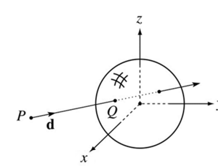
Finally, when shading a point inside of an object, apply k_t to the ambient component, since that "ambient light" had to pass through the object to get there in the first place.



Intersecting rays with spheres

Now we've done everything except figure out what that "scene.*intersect*(*P*, **d**)" function does.

Mostly, it calls each object to find out the t value at which the ray intersects the object. Let's start with intersecting spheres…



Given:

- The coordinates of a point along a ray passing through *P* in the direction **d** are:

$$x = P_x + td_x$$
$$y = P_y + td_y$$
$$z = P_z + td_z$$

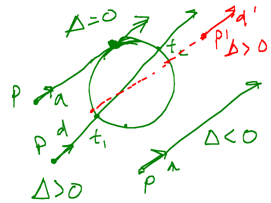
$r(t) = P + td$

- A unit sphere *S* centered at the origin defined by the equation:

$x^2 + y^2 + z^2 = 1$ $\qquad (P_x+td_x)^2 + (P_y+td_y)^2 + (P_z+td_z)^2 = 1$

Find: The *t* at which the ray intersects *S*.

Intersecting rays with spheres



Solution by substitution:

$$x^2 + y^2 + z^2 - 1 = 0$$
$$(P_x + td_x)^2 + (P_y + td_y)^2 + (P_z + td_z)^2 - 1 = 0$$
$$at^2 + bt + c = 0$$

$\Delta = 0 \Rightarrow$ no $\cap$

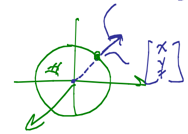
where

$$a = d_x^2 + d_y^2 + d_z^2$$
$$b = 2(P_x d_x + P_y d_y + P_z d_z)$$
$$c = P_x^2 + P_y^2 + P_z^2 - 1$$

Q: What are the solutions of the quadratic equation in *t* and what do they mean?

$$t = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \qquad \Delta = b^2 - 4ac$$

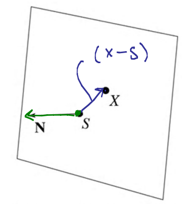
discriminant



Q: What is the normal to the sphere at a point (*x,y,z*) on the sphere?

$$N = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Ray-plane intersection



$$X = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Next, we will considering intersecting a ray with a plane.

To do this, we first need to define the plane equation.

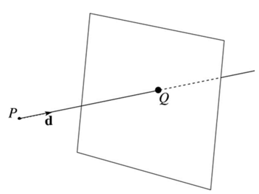
Given a point *S* on a plane with normal **N**, how would we determine if a point *X* is on the plane?

$$N \cdot (X - S) = 0$$
$$N \cdot X - N \cdot S = 0$$
$$N \cdot X = N \cdot S$$
$$N_x x + N_y y + N_z z = K$$

$$ax + by + cz = k$$

This is the plane equation!

Ray-plane intersection (cont'd)



Now consider a ray intersecting a plane. The plane has equation:

$$N_x x + N_y y + N_z z = K \qquad N \cdot X = K$$

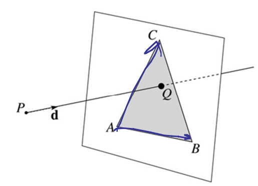
We can solve for the intersection parameter (and thus the point):

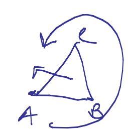
$$r(t) = P + td$$
$$N \cdot (P + td) = K$$
$$N \cdot P + N \cdot td = K$$
$$N \cdot P + tN \cdot d = K$$
$$tN \cdot d = K - N \cdot P \Rightarrow t = \frac{K - N \cdot P}{N \cdot d}$$

if $N \cdot d = 0 \Rightarrow$ "no $\cap$"

d // plane

Ray-triangle intersection





To intersect with a triangle, we first solve for the equation of its supporting plane.

How might we compute the (un-normalized) normal?

$$N = (B - A) \times (C - A)$$

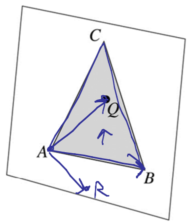
Given this normal, how would we compute *k*?

$$N \cdot X = K = N \cdot A = N \cdot B = N \cdot C \quad \text{(pick one)}$$

Using these coefficients, we can solve for *Q*. Now, we need to decide if *Q* is inside or outside of the triangle.

3D inside-outside test

One way to do this "inside-outside test," is to see if *Q* lies on the left side of each edge as we move counterclockwise around the triangle.

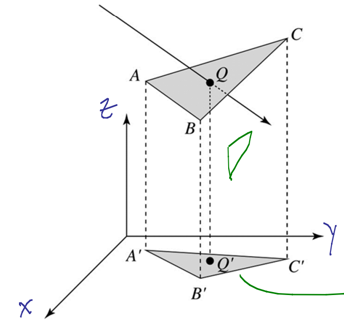


How might we use cross products to do this?

$$\left.\begin{array}{l}[(B-A)\times(Q-A)]\cdot N \geq 0 \\ [(C-B)\times(Q-B)]\cdot N \geq 0 \\ [(A-C)\times(Q-C)]\cdot N \geq 0\end{array}\right\} \Rightarrow Q \text{ is inside } \Delta_{ABC}$$

2D inside-outside test

Without loss of generality, we can make this determination after projecting down a dimension:



$Q = \alpha A + \beta B + \gamma C$

$Q' = \alpha A' + \beta B' + \gamma C'$

If *Q′* is inside of *A′B′C′*, then *Q* is inside of *ABC*.

Why is this projection desirable? Cheaper cross products

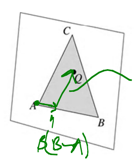
Which axis should you "project away"?

axis for which normal component is largest

Barycentric coordinates

As we'll see in a moment, it is often useful to represent *Q* as an **affine combination** of *A*, *B*, and *C*:

$$Q = \alpha A + \beta B + \gamma C$$



where:

$$\alpha + \beta + \gamma = 1$$

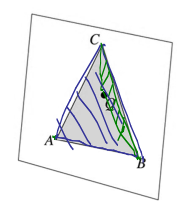
We call α, β, and γ, the **barycentric coordinates** of *Q* with respect to *A*, *B*, and *C*.

$$Q = A + \beta(B-A) + \gamma(C-A)$$
$$= (1-\beta-\gamma)A + \beta B + \gamma C$$
$$= \alpha A + \beta B + \gamma C$$
$$\alpha = 1-\beta-\gamma \Rightarrow \alpha+\beta+\gamma = 1$$

Computing barycentric coordinates

Given a point *Q* that is inside of triangle *ABC*, we can solve for *Q*'s barycentric coordinates in a simple way:

$$\alpha = \frac{\text{Area}(QBC)}{\text{Area}(ABC)} \quad \beta = \frac{\text{Area}(AQC)}{\text{Area}(ABC)} \quad \gamma = \frac{\text{Area}(ABQ)}{\text{Area}(ABC)}$$



How can cross products help here?

see HW

In the end, these calculations can be performed in the 2D projection as well!

Interpolating vertex properties

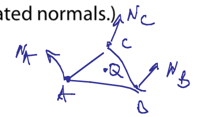
The barycentric coordinates can also be used to interpolate vertex properties such as:

- material properties
- texture coordinates
- normals

For example:

$$k_d(Q) = \alpha k_d(A) + \beta k_d(B) + \gamma k_d(C)$$

Interpolating normals, known as **Phong *interpolation***, gives triangle meshes a smooth shading appearance. (Note: don't forget to normalize interpolated normals.)



α, β, γ for Q

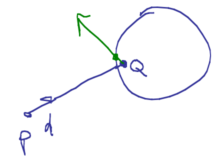
$$N_Q = \alpha N_A + \beta N_B + \gamma N_C$$

$$\hat{N}_Q = \frac{N_Q}{\|N_Q\|}$$

Epsilons

Due to finite precision arithmetic, we do not always get the exact intersection at a surface.

Q: What kinds of problems might this cause?



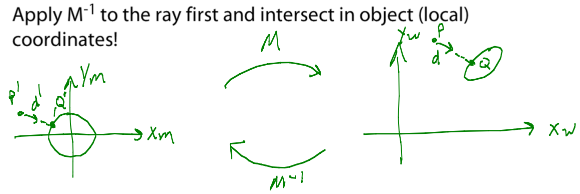
Q: How might we resolve this?

$t <$ RAY_EPSILON $\Rightarrow$ no $\cap$

Intersecting with xformed geometry

In general, objects will be placed using transformations. What if the object being intersected were transformed by a matrix M?

Apply M^{-1} to the ray first and intersect in object (local) coordinates!



$$M^{-1}(Q = p + t_n d)$$

$$M^{-1}Q = M^{-1}(p + t_n d)$$

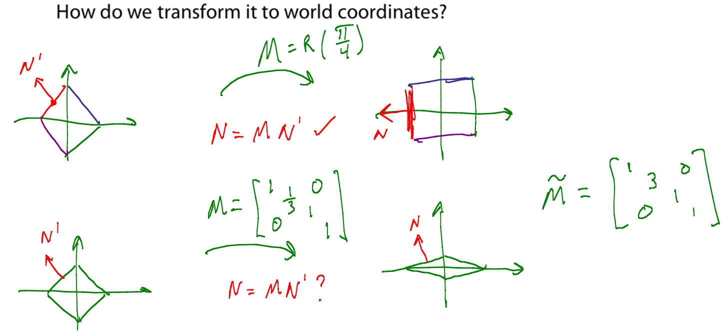
$$M^{-1}Q = M^{-1}p + M^{-1}t_n d$$

$$M^{-1}Q = M^{-1}p + t_n M^{-1}d$$

$$Q' = p' + t_n d'$$

Intersecting with xformed geometry

The intersected normal is in object (local) coordinates. How do we transform it to world coordinates?

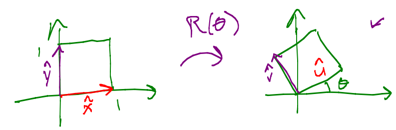


$$M = \left[\begin{array}{c|c} A_{3\times3} & t_{3\times1} \\ \hline 0\ 0\ 0 & 1 \end{array}\right] \rightarrow \tilde{M} = \left[\begin{array}{c|c} A^{-T}_{3\times3} & t \\ \hline 0\ 0\ 0 & 1 \end{array}\right] \qquad N = \tilde{M} N'$$

Summary

What to take home from this lecture:

- The meanings of all the boldfaced terms.
- Enough to implement basic recursive ray tracing.
- How reflection and transmission directions are computed.
- How ray-object intersection tests are performed on spheres, planes, and triangles
- How barycentric coordinates within triangles are computed
- How ray epsilons are used.



$$R = \begin{bmatrix} \hat{u} & \hat{v} \end{bmatrix}$$

$$R = \begin{bmatrix} \vdots & \vdots & \vdots \\ \hat{u} & \hat{v} & \hat{w} \\ \vdots & \vdots & \vdots \end{bmatrix}$$

$$R^T R = \begin{bmatrix} \cdots \hat{u}^T \cdots \\ \cdots \hat{v}^T \cdots \\ \cdots \hat{w}^T \cdots \end{bmatrix} \begin{bmatrix} \hat{u} & \hat{v} & \hat{w} \end{bmatrix} = \begin{bmatrix} \hat{u}^T\hat{u} & \hat{u}^T\hat{v} & \hat{u}^T\hat{w} \\ \hat{v}^T\hat{u} & \hat{v}^T\hat{v} & \hat{v}^T\hat{w} \\ \hat{w}^T\hat{u} & \hat{w}^T\hat{v} & \hat{w}^T\hat{w} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I$$

$$R^T R = I$$

$$R^T = R^{-1}$$