# Ray Sphere Intersection

November 19, 2022

## 1 Analytic Solution

Remember that a ray can be expressed using the following function: $O + tD$ (equation 1) where $O$ is a point and corresponds to the origin of the ray, $D$ is a vector and corresponds to the direction of the ray, and $t$ is a parameter of the function. By varying $t$ (which can be either positive or negative) we can compute any point on the line defined by the ray origin and direction. When $t$ is greater than 0, then the point on the ray is in "front" of the ray's origin. When $t$ is negative, the point is behind the ray's origin. When $t$ is exactly 0, the point and the ray's origin are the same.

The idea behind solving the ray-sphere intersection test, is that spheres too can be defined using an algebraic form. The equation for a sphere is:

$$x^2 + y^2 + z^2 = R^2$$

Where x, y and z are the coordinates of a cartesian point and $R$ is the radius of a sphere centred at the origin (will see later how to change the equation so that it works with spheres which are not centred at the origin). It says that there is a set of points for which the above equation is true. This set of points defines the surface of a sphere which is centred at the origin and has radius $R$. Or more simply, if we consider that x, y, z are the coordinates of point P, we can write (equation 2):

$$P^2 - R^2 = 0$$

This equation is typical of what we call in Mathematics and CG an implicit function and a sphere expressed in this form is also called an implicit shape or surface. Implicit shapes are shapes which can be defined not in terms of polygons connected to each other for instance (which is the type of geometry you might be familiar with if you have modelled object in a 3D application such as Maya or Blender) but simply in terms of equations. Many shapes (often quite simple though) can be defined in terms of a function (cube, cone, sphere, etc.). However simple, these shapes can be combined together to create more complex forms. This is the idea behind modeling geometry using blobs for instance (blobby surfaces are also called metaballs). But before we got too far off course here, let's get back to the ray-sphere intersection test (check the advanced section for a lesson on Implicit Modeling).

All we need to do now, is to substitute equation 1 in equation 2 that is, to replace P in equation 2 with the equation of the ray (remember that O+tD defines all points along the ray):

$$|O + tD|^2 - R^2 = 0$$

When we develop this equation we get (equation 3):

$$O^2 + (Dt)^2 + 2ODt - R^2 = O^2 + D^2t^2 + 2ODt - R^2 = 0$$

which in itself is an equation of the form (equation 4):

$$f(x) = ax^2 + bx + c$$

with $a = D^2$, b=2OD and $c = O^2 - R^2$ (remember that x in equation 4 corresponds to $t$ in equation 3 which is the unknown). Being able to re-write equation 3 into equation 4 is important because equation 4 is known as a quadratic function. It is a function for which the roots (when x takes a value for which f(x) = 0) can easily be found using the following equations (equation 5):

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

Note the +/- sign in the formula. The first root uses the sign + and the second root uses the sign -. The letter $\Delta$ (Greek letter delta) is called the discriminant. The sign of the discriminant indicates whether there is two, one or no root to the equation.:

when $\Delta > 0$ there is two roots which can be computed with:

$$\frac{-b+\sqrt{\Delta}}{2a} \quad and \quad \frac{-b-\sqrt{\Delta}}{2a}$$

In that case, the ray intersects the sphere in two places (at $t_0$ and $t_1$). when $\Delta$ = 0 there is one root which can be computed with:

$$-\frac{b}{2a}$$

The ray intersects the sphere in one place only ($t_0$=$t_1$). when $\Delta < 0$, there is not root at (which means that the ray doesn't intersect the sphere). Since we have a, b and c, we can easily compute these equations to get the values for $t$ which corresponds to the two intersection points of the ray with the sphere ($t_0$ and $t_1$ in figure 1). Note that the root values can be negative which means that the ray intersects the sphere but behind the origin. One of the roots can be negative and the other positive which means that the origin of the ray is inside the sphere. There also might be no solution to the quadratic equations which means that the ray doesn't intersect the sphere at all (no intersection between the ray and the sphere).

Since we have a, b and c, we can easily compute these equations to get the values for $t$ which corresponds to the two intersection points of the ray with the sphere ($t_0$ and $t_1$ in figure 1). Note that the root values can be negative which means that the ray intersects the sphere but behind the origin. One of the roots can be negative and the other positive which means that the origin of the ray is inside the sphere. There also might be no solution to the quadratic equations which means that the ray doesn't intersect the sphere at all (no intersection between the ray and the sphere).

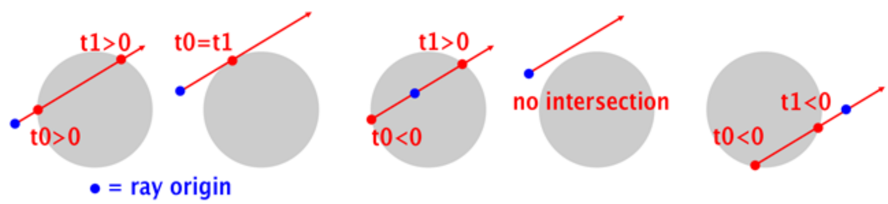


Figure 1: When a ray is tested for an intersection with a sphere, several cases might be considered. It is important to properly deal with cases where the intersection is behind the origin of the ray (spheres 3 and 5). These intersections might sometimes be undesirable.