Alhazen's Problem: Reflection Point on a Sphere

David Eberly, Geometric Tools, Redmond WA 98052
https://www.geometrictools.com/

Created: February 1, 2008
Last Modified: May 26, 2022

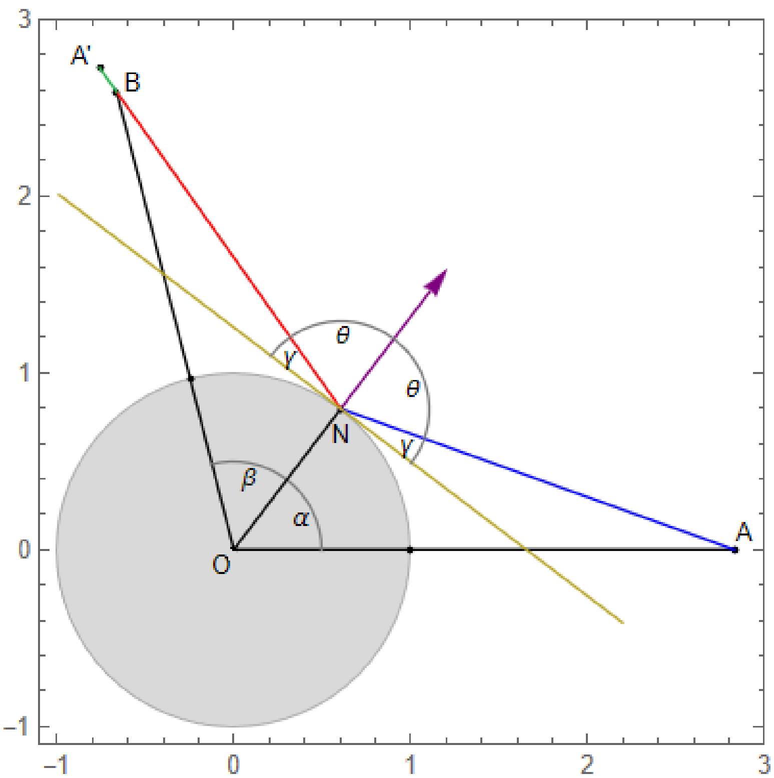
Contents

1 Introduction

A point light is located at position $\boldsymbol{A}$, which is more than one unit of distance from the origin. A sphere of radius 1 and centered at the origin $\boldsymbol{O}$ will reflect light rays from the point light. A point $\boldsymbol{B}$ outside the sphere potentially receives a reflected ray of light. If it does, we wish to compute the point $\boldsymbol{N}$ on the sphere at which the light ray is reflected to reach the point $\boldsymbol{B}$. The problem is known as *Alhazen's problem* [2], a mathematical problem in geometric optics first formulated by Ptolemy in 150 A.D. It was phrased solely in terms of trigonometric functions. Using linear algebra and basic algebra, the problem can be formulated in a manner that requires computing roots to a quartic polynomial [1].[1]

Figure 1 illustrates the problem, effectively a 2-dimensional problem by viewing it in the plane containing $\boldsymbol{O}$, $\boldsymbol{A}$ and $\boldsymbol{B}$.

Figure 1. The light is located at $\boldsymbol{A}$. A point that receives the reflected light is located at $\boldsymbol{B}$. The unit-length point $\boldsymbol{N}$ is the point where a light beam from $\boldsymbol{A}$ is reflected and received by $\boldsymbol{B}$. The gold-colored line is tangent to the circle at $\boldsymbol{N}$.



The point $\boldsymbol{B}$ will receive a reflected light ray as long as it is outside the sphere and not in the shadow of the

[1] The previous version of the document was written based on a request from someone to compute the point $\boldsymbol{N}$. I was unaware that the problem had a name and that there was a publication that showed how to compute the quartic polynomial. The title of [1] is "A deceptively easy problem", which I consider to be misleading. The algebra used to construct the quartic polynomial is easy and straightforward; there is nothing deceptive about it.

sphere, as illustrated in Figure 1. The point $\boldsymbol{A}'$ is the reflection of $\boldsymbol{A}$ through the ray whose origin is the sphere center and whose direction is $\boldsymbol{N}$.

2 Trigonometric Formulation

The trigonometric formulation is based on the Law of sines for a triangle. Let a triangle have angles ϕ_0, ϕ_1 and ϕ_2. The sides opposite the angles have lengths ℓ_0, ℓ_1 and ℓ_2. The Law of sines is

$$\frac{\sin(\phi_0)}{\ell_0} = \frac{\sin(\phi_1)}{\ell_1} = \frac{\sin(\phi_2)}{\ell_2} \tag{1}$$

Consider the triangle in figure 1 with vertices $\boldsymbol{A}$, $\boldsymbol{O}$ and $\boldsymbol{N}$. The angle at $\boldsymbol{O}$ is α and has opposite side of length $|\boldsymbol{A}-\boldsymbol{N}|$. The angle at $\boldsymbol{N}$ is $\pi/2+\gamma$ and has opposite side length of $|\boldsymbol{A}|$. The angle at $\boldsymbol{A}$ is $\pi/2-\alpha-\gamma$ and has opposite side length of 1, the radius of the sphere. Applying equation (1) only at vertices $\boldsymbol{N}$ and $\boldsymbol{A}$,

$$\frac{\cos(\gamma)}{|\boldsymbol{A}|} = \frac{\sin(\pi/2+\gamma)}{|\boldsymbol{A}|} = \frac{\sin(\pi/2-\alpha-\gamma)}{1} = \cos(\alpha)\cos(\gamma) - \sin(\alpha)\sin(\gamma) \tag{2}$$

The middle equality comes from the Law of sines. The first and last equalities are based on trigonometric identities. Dividing by $\cos\gamma$, we may solve this for

$$\tan(\gamma) = \frac{\cos(\alpha) - 1/|\boldsymbol{A}|}{\sin(\alpha)} \tag{3}$$

A similar argument applies to the triangle in figure 1 with vertices $\boldsymbol{B}$, $\boldsymbol{O}$ and $\boldsymbol{N}$ to produce

$$\tan(\gamma) = \frac{\cos(\beta) - 1/|\boldsymbol{B}|}{\sin(\beta)} \tag{4}$$

The vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ are known quantities, so the angle $\phi = \alpha+\beta$ between them is known and is determined by the dot product $\cos\phi = (\boldsymbol{A}/|\boldsymbol{A}|)\cdot(\boldsymbol{B}/|\boldsymbol{B}|)$. The difference of equations (3) and (4) is a function only of $\alpha \in [0,\phi]$,

$$F(\alpha) = \frac{\cos(\alpha) - 1/|\boldsymbol{A}|}{\sin(\alpha)} - \frac{\cos(\phi-\alpha) - 1/|\boldsymbol{B}|}{\sin(\phi-\alpha)} \tag{5}$$

A numerical root finder may be used to solve $F(\alpha) = 0$ for $\alpha \in [0,\phi]$. Notice that the numerator of the first term on the right of equation (5) when $\alpha = 0$ is $1 - 1/|\boldsymbol{A}| > 0$. The numerator of the second term on the right of equation (5) at $\alpha = \phi$ is $1 - 1/|\boldsymbol{B}| > 0$. The two conditions imply

$$F(0) = \lim_{\alpha\to 0^+} F(\alpha) = +\infty, \quad F(\phi) = \lim_{\alpha\to\phi^-} F(\alpha) = -\infty \tag{6}$$

Geometrically there must be a unique solution $\boldsymbol{N}$, which means there is a unique α for which F is zero. It is sufficient to use a bisection algorithm to estimate the root of $F(\alpha)$. Listing 1 contains pseudocode for this.

Listing 1. Pseudocode for computing $\boldsymbol{N}$ by solving $F(\alpha) = 0$ with $\alpha \in [0, \phi]$. The algorithm uses bisection.

```
Real F(Real alpha, Real invLengthA, Real invLengthB, Real phi)
{
    Real term0 = (cos(alpha) - invLengthA) / sin(alpha);
    Real beta = phi - alpha;
    Real term1 = (cos(beta) - invLengthB) / sin(beta);
    Real result = term0 - term1;
    return result;
}

// Assert: A and B have lengths larger than 1. B is outside the cone formed by A and its tangents to the sphere or B is
// inside the cone and is in front of the sphere (visible to A). The return value is N.
Vector2<Real> Solver(Vector2<Real> A, Vector2<Real> B, uint maxIterations)
{
    Real invLengthA = 1 / Length(A);
    Real invLengthB = 1 / Length(B);
    Real phi = acos(Dot(A * invLengthA, B * invLengthB));

    // F(alpha_min) = +infinity and F(alpha_max) = -infinity. However, all that is needed for bisection are the signs of F at the endpoints
    // of the alpha-intervals.
    Real alphaMin = 0, alphaMax = phi, alpha0 = alphaMin, alpha1 = alphaMax, alphaRoot = 0;
    int signMin = +1, signMax = -1;
    for (uint iteration = 0; iteration <= maxIterations; ++iteration)
    {
        Real alphaMid = (alpha0 + alpha1) / 2;
        Real fMid = F(alphaMid, invLengthA, invLengthB, phi);

        // If signMid is 0, we have found the root exactly (low probability). If alphaMid equals alpha0 or alphaMid
        // equals alpha1, the interval endpoints are consecutive floating-point numbers, so subdividing is no longer
        // possible and we have the best estimate of the root.
        int signMid = (fMid > 0 ? +1 : (fMid < 0 ? -1 : 0));
        if (signMid == 0 || alphaMid == alpha0 || alphaMid == alpha1)
        {
            alphaRoot = alphaMid;
            break;
        }

        // Update the correct endpoint to the midpoint.
        if (signMid == signMin)
        {
            alpha0 = alphaMid;
        }
        else // signMid == signMax
        {
            alpha1 = alphaMid;
        }
    }

    Vector2<Real> N = { cos(alphaRoot), sin(alphaRoot) };
    return N;
}
```

3 Quartic Polynomial Formulation

Figure 1 shows the reflection $\boldsymbol{A}'$ of the point $\boldsymbol{A}$ through the normal $\boldsymbol{N}$. We can write $\boldsymbol{A} = s\boldsymbol{N} + t\boldsymbol{N}^{\perp}$, where $\boldsymbol{N}^{\perp}$ is a unit-length vector perpendicular to $\boldsymbol{N}$. The reflection is $\boldsymbol{A}' = s\boldsymbol{N} - t\boldsymbol{N}^{\perp}$, which implies

$$\boldsymbol{A}' = 2(\boldsymbol{N} \cdot \boldsymbol{A})\boldsymbol{N} - \boldsymbol{A} \tag{7}$$

The assumption in this section is that $\boldsymbol{A}$ and $\boldsymbol{B}$ are not parallel vectors. For if they were parallel, the point of reflection is $\boldsymbol{N} = \boldsymbol{A}/|\boldsymbol{A}|$; that is, the light is reflected in the opposite direction to reach $\boldsymbol{B}$. Our assumption has the consequence that $\boldsymbol{A} \times \boldsymbol{B} \neq \boldsymbol{0}$; that is, non-parallel vectors have a nonzero cross product.

Figure 1 shows that $\boldsymbol{N}$ must bisect the angle between $\boldsymbol{A} - \boldsymbol{N}$ and $\boldsymbol{B} - \boldsymbol{N}$. We can represent

$$\boldsymbol{N} = x\boldsymbol{A} + y\boldsymbol{B} \tag{8}$$

for some scalars $x > 0$ and $y > 0$. Observe that

$$\boldsymbol{A} \times \boldsymbol{N} = y\boldsymbol{A} \times \boldsymbol{B}, \quad \boldsymbol{N} \times \boldsymbol{B} = x\boldsymbol{A} \times \boldsymbol{B}, \quad \boldsymbol{N} \cdot \boldsymbol{A} = x\boldsymbol{A} \cdot \boldsymbol{A} + y\boldsymbol{A} \cdot \boldsymbol{B} \tag{9}$$

Because $\boldsymbol{N}$ is a unit-length vector, we also know that

$$1 = \boldsymbol{N} \cdot \boldsymbol{N} = x^2\boldsymbol{A} \cdot \boldsymbol{A} + 2xy\boldsymbol{A} \cdot \boldsymbol{B} + y^2\boldsymbol{B} \cdot \boldsymbol{B} \tag{10}$$

$\boldsymbol{N}$ must be chosen so that the vectors $\boldsymbol{B} - \boldsymbol{N}$ and $\boldsymbol{A}' - \boldsymbol{N}$ are parallel. Their cross product must be the zero vector,

$$\begin{aligned} \boldsymbol{0} &= (\boldsymbol{B} - \boldsymbol{N}) \times (\boldsymbol{A}' - \boldsymbol{N}) \\ &= (\boldsymbol{B} - \boldsymbol{N}) \times [(2\boldsymbol{N} \cdot \boldsymbol{A} - 1)\boldsymbol{N} - \boldsymbol{A}] \\ &= (2\boldsymbol{N} \cdot \boldsymbol{A} - 1)\boldsymbol{B} \times \boldsymbol{N} - \boldsymbol{B} \times \boldsymbol{A} + \boldsymbol{N} \times \boldsymbol{A} \\ &= [(2\boldsymbol{N} \cdot \boldsymbol{A} - 1)x - 1 + y]\boldsymbol{B} \times \boldsymbol{A} \\ &= [(2x\boldsymbol{A} \cdot \boldsymbol{A} + 2y\boldsymbol{A} \cdot \boldsymbol{B} - 1)x - 1 + y]\boldsymbol{B} \times \boldsymbol{A} \end{aligned} \tag{11}$$

where we have used equations (7), (8), and (9). By assumption, $\boldsymbol{B} \times \boldsymbol{A} \neq \boldsymbol{0}$, so equation (11) implies

$$(2x\boldsymbol{A} \cdot \boldsymbol{A} + 2y\boldsymbol{A} \cdot \boldsymbol{B} - 1)x - 1 + y = 0 \tag{12}$$

Equations (10) and (12) are two quadratic equations in two unknowns x and y,

$$p(x, y) = ux^2 + 2vxy + wy^2 - 1 = 0, \quad q(x, y) = 2ux^2 + 2vxy - x + y - 1 = 0 \tag{13}$$

where $u = \boldsymbol{A} \cdot \boldsymbol{A}$, $v = \boldsymbol{A} \cdot \boldsymbol{B}$ and $w = \boldsymbol{B} \cdot \boldsymbol{B}$. We may solve $q(x, y) = 0$ for y in terms of x,

$$y = \frac{1 + x - 2ux^2}{1 + 2vx} \tag{14}$$

Substituting into the equation $p(x, y) = 0$, we have

$$\frac{4u(uw - v^2)x^4 + 4(v^2 - uw)x^3 + (u + 2v + w - 4uw)x^2 + 2(w - v)x + (w - 1)}{(1 + 2vx)^2} = 0 \tag{15}$$

The numerator of equation (15) is the quartic polynomial,

$$r(x) = 4u(uw - v^2)x^4 + 4(v^2 - uw)x^3 + (u + 2v + w - 4uw)x^2 + 2(w - v)x + (w - 1) \tag{16}$$

Observe that

$$uw - v^2 = (\boldsymbol{A} \cdot \boldsymbol{A})(\boldsymbol{B} \cdot \boldsymbol{B}) - (\boldsymbol{A} \cdot \boldsymbol{B})^2 = |\boldsymbol{A} \times \boldsymbol{B}|^2 \neq 0 \tag{17}$$

so the coefficient of x^4 is not zero, which means $r(x)$ really has degree 4.

Now compute the real-valued roots of $r(x) = 0$. For each root $\bar{x} > 0$ use equation (14) to compute $\bar{y} = (1 + \bar{x} - 2u\bar{x}^2)/(1 + 2v\bar{x})$. Of all the pairs $(\bar{x}, \bar{y})$, select that pair for which $\bar{x} > 0$ and $\bar{y} > 0$. The point of reflection is $\boldsymbol{N} = \bar{x}\boldsymbol{A} + \bar{y}\boldsymbol{B}$. Listing 2 contains pseudocode for the algorithm.

Listing 2. Pseudocode for computing $\boldsymbol{N}$ by computing the appropriate root of a quartic polynomial.

```
// Assert: A and B have lengths larger than 1. B is outside the cone formed by A and its tangents to the sphere or B is
// inside the cone and is in front of the sphere (visible to A). The return value is N.
Vector2<Real> Solver(Vector2<Real> A, Vector2<Real> B)
{
    Real u = A[0] * A[0] + A[1] * A[1];
    Real v = A[0] * B[0] + A[1] * B[1];
    Real w = B[0] * B[0] + B[1] * B[1];

    Real r0 = w - 1;
    Real r1 = 2 * (w - v);
    Real r2 = u + 2 * v + w - 4 * u * w;
    Real r3 = 4 * (v * v - u * w);
    Real r4 = 4 * u * (u * w - v * v);
    int numRoots = 0;
    array<Real, 4> roots;
    SolveQuartic(r0, r1, r2, r3, r4, numRoots, roots);
    Real x, y;
    for (int i = 0; i < numRoots; ++i)
    {
        x = roots[0];
        if (x > 0)
        {
            y = (1 + x - 2 * u * x * x) / (1 + 2 * v * x);
            if (y > 0)
            {
                break;
            }
        }
    }

    Vector2<Real> N = x * A + y * B;

    // If you want to know angles show in figure 1, return these to the caller.
    Real alpha = atan2(N[1], N[0]);
    Real phi = acos(Dot(A / Length(A), B / Length(B)));
    Real beta = phi - alpha;

    // γ0 and γ1 are theoretically the same, but numerically there may be some floating-point rounding errors that cause
    // them to be slightly different.
    Real gamma0 = atan((cos(alpha) - 1 / Length(A)) / sin(alpha));
    Real gamma1 = atan((cos(beta) - 1 / Length(B)) / sin(beta));

    return N;
}
```

The function SolveQuartic should be of good quality and return only the real-valued roots of the polynomial $r(x) = r_0 + r_1 x + r_2 x^2 + r_3 x^3 + r_4 x^4$.

References

[1] Jack M. Elkin. A deceptively easy problem. *The Mathematics Teacher*, 58(3):194–199, 1965. https://www.jstor.org/stable/27968003.

[2] Wikipedia. Alhazen's problem. https://en.wikipedia.org/wiki/Alhazen's_problem. accessed May 24, 2022.