

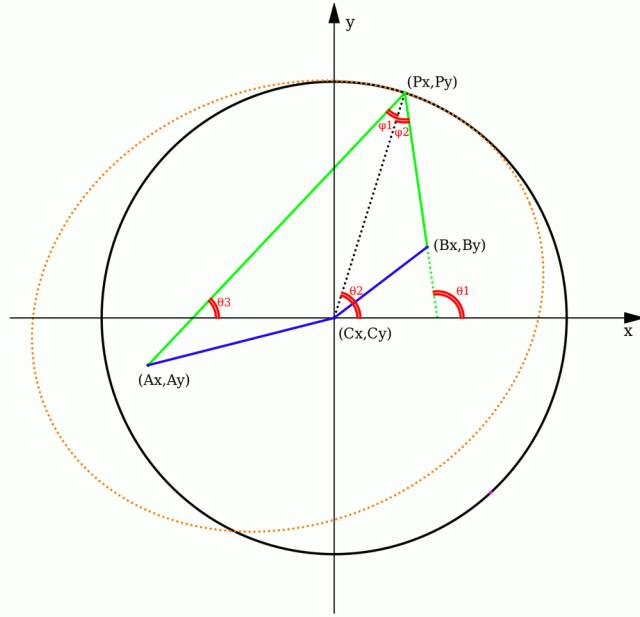
ALHAZEN'S PROBLEM

The problem was first formulated by Ptolemy in 150 AD, and was named after the Arab scholar Alhazen, who discussed it in his work on optics and gave a solution. [1]

The problem asks that, given two points A and B and a spherical mirror with radius r, centered at point C, find the point on mirror that when a ray of light originating at A will be reflected in such a way that it will pass through B.

It is sometimes known as "Alhazen's Billiards problem" because it may be formulated as finding the point on the boundary of a circular billiards table at which the cue ball must be aimed, if it is to hit the second ball after one bounce off the cushion.

This problem is equivalent to the problem that given the 2 foci of an ellipse find the ellipse that it will be tangent to a given circle. Generally, there are up to 4 such ellipses depending on the relative positions of the points and circle.



I came across this problem, some years ago, in the context of mechanics: Imagine that we have an arm, free to rotate around point C, and two fixed points A and B. If we pass a rope through a hole at the end of the arm and tighten the rope at A and B what is the final position of the arm? I solved it at the time with successive approximations starting with a random point. A few years later I discovered accidentally that this was Alhazen's problem.

There are several ways to solve this problem, as one can easily verify by searching the web. Some people use the fact that A and B are the foci of an ellipse, tangent to the circle, and proceed to find the solution. However I will follow the solution in [2], also revised in [6]. Another formulation was given by N.Kollas [7] who gives an elegant formula with one unknown. I originally followed this, but the unknown is a tangent and in order to pick the right solution, up to eight points had to be checked. I present this solution in the appendix II.

Suppose that we have two points, A and B and one circle with center C and radius r. We translate the whole set so that the circle center is at (0,0). The requirement is to find point P, which will reflect a ray from any point to the other one.

We denote the angle of incidence by ϕ_1 and the angle of reflection by ϕ_2 and furthermore introduce the angles θ_1, θ_2 and θ_3 between the x-axis and the lines PB, PC and PA and RP2 respectively. We then have:

$$\phi_1 = \theta_2 - \theta_3 \quad \text{and} \quad \phi_2 = \theta_1 - \theta_2 \quad \text{and from the law of reflection} \quad \phi_1 = \phi_2$$

By plugging $\tan(\theta_2) = \frac{P_y}{P_x}$, $\tan(\theta_3) = \frac{P_y - A_y}{P_x - A_x}$, $\tan(\theta_1) = \frac{P_y - B_y}{P_x - B_x}$ into the equation $\tan(\phi_1) = \tan(\phi_2)$

we obtain, after some algebra, $\frac{P_x \cdot A_y - P_y \cdot A_x}{P_x^2 + P_y^2 - P_x \cdot A_x - P_y \cdot A_y} = \frac{P_x \cdot B_y - P_y \cdot B_x}{P_x^2 + P_y^2 - P_x \cdot B_x - P_y \cdot B_y}$ This is a third degree equation for P_x

and P_y . Since we are only interested only in reflection points on the circle, we can use the substitution:

$$r^2 = P_x^2 + P_y^2 \quad \text{so the formula takes the form} \quad \frac{P_x \cdot A_y - P_y \cdot A_x}{r^2 - P_x \cdot A_x - P_y \cdot A_y} = \frac{P_x \cdot B_y - P_y \cdot B_x}{r^2 - P_x \cdot B_x - P_y \cdot B_y}$$

Reduction to a standard form yields:

$$A \cdot x^2 + 2 \cdot C \cdot x \cdot y - A \cdot y^2 + r^2 \cdot (2 \cdot D \cdot x + 2 \cdot E \cdot y) = 0 \quad \text{or} \quad A \cdot x^2 + 2 \cdot C \cdot x \cdot y - A \cdot y^2 + 2 \cdot r^2 \cdot D \cdot x + 2 \cdot r^2 \cdot E \cdot y = 0 \quad (12)$$

where (x,y) are the coordinates of P (P_x, P_y) and $A = A_x \cdot B_y - B_x \cdot B_y$ $C = A_y \cdot B_y - A_x \cdot B_x$ $D = -0.5 \cdot (A_y + B_y)$

$$E = 0.5 \cdot (A_x + B_x)$$

Equation 12 can be transformed to the general conic form $A_1 \cdot x^2 + B_1 \cdot x \cdot y + C_1 \cdot y^2 + D_1 \cdot x + E_1 \cdot y + F_1 = 0$ by doing the replacements: $A_1 = A$, $B_1 = 2 \cdot C$, $C_1 = -A$, $D_1 = 2 \cdot r^2 \cdot D$, $E_1 = 2 \cdot r^2 \cdot E$ and $F_1 = 0$

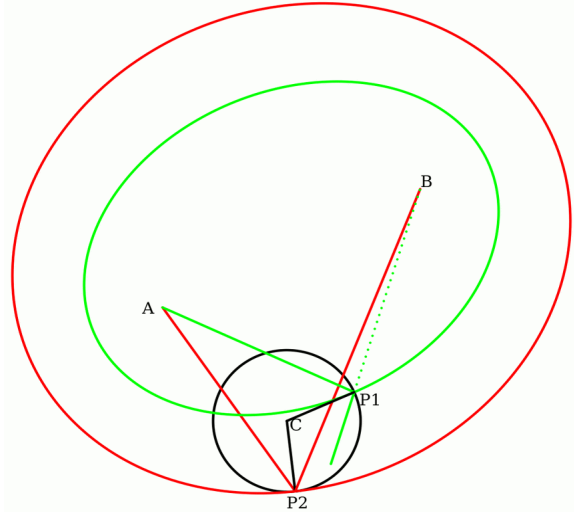
Since $B_1^2 - 4 \cdot A_1 \cdot C_1 = (2 \cdot C)^2 + 4 \cdot A^2 > 0$ this conic is a hyperbola and since $A_1 = -C_1$ it is a rectangular one.

Equation (12) shows that the points having equal reflection angles with points A and B lie on this hyperbola, so we conclude that the reflection point P is the intersection of this hyperbola and the circle.

The circle with radius r and center coordinates (0,0) can be represented in general conic form as

$A_2 \cdot x^2 + B_2 \cdot x \cdot y + C_2 \cdot y^2 + D_2 \cdot x + E_2 \cdot y + F_2 = 0$ where $A_2 = C_2 = r^2$, $B_2 = D_2 = E_2 = 0$, $F_2 = -r^4$ (this can be proved easily by expanding the general circle equation $(x - c_x)^2 + (y - c_y)^2 = r^2$ centered at (c_x, c_y)).

This procedure produces up to 4 real solutions, so there are up to 4 points on the circle on which the ray is reflected with equal angles to the radius at the reflection point. However, when at least one point is outside the circle, some rays are not reflected to the other point but they are reflected to the interior of the circle and the second point only lies on the extension of the reflected ray. At these reflection points the ray appears to use the radius as a reflection line and not the tangent at the point. As can be seen in the adjacent picture, point P1 is such a "false" reflection point. At P1 the tangents to the circle and to resulting ellipse are perpendicular to each other. We use this fact to weed out these false points.



In order to select the right reflection point, we employ the following procedure. For each point found, its distance to the two points A and B is calculated. The sum of these distances, $AP + PB$ is the length of the major axis α of the

ellipse. The minor semi-axis β is calculated with the formula $\beta^2 = \left(\frac{\alpha}{2}\right)^2 - \left(\frac{rC}{2}\right)^2$ where rC is the distance AB , so the ellipse is totally defined. If the slope of the tangents to this particular point is the same for ellipse and the circle the point is selected as a proper solution.

For any conic $Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$ the slope, m , of the tangent at any point (x_f, y_f) is given by its first derivative as $m = -\frac{(2 \cdot A \cdot x_f + B \cdot y_f + D)}{(B \cdot x_f + 2 \cdot C \cdot y_f + E)}$. If this value is not the same for both figures, this solution is not taken into account. In such case the product of the slopes is equal to -1, meaning that ellipse and circle are orthogonal as can easily be verified geometrically.

This program produces a drawing in DXF format. This can be opened with any convenient CAD tool such as LibreOffice Cad and Qcad, both of which are free software. In the drawing, the points A, B and C are marked with letters and the lines to each reflection point are marked with the same color. The corresponding ellipses are marked with a duller color for each point.

The accuracy of the calculations deteriorates when one or both points are near to, or on, the circle. When A, B and C are collinear and A or B lies on the circle, a degenerate ellipse is found with b axis zero or very small.

For the solution of quartic equation, the NBS method is applied together with a trigonometric method for the cubic equation (compiled after the instructions given by David Wolters in <http://www.quarticequations.com/>). Most analytic geometry functions were compiled from <http://www.ambrsoft.com/TrigoCalc/menu/menuTrigoCalc.htm>

For the compilation of DXF files, the DXF fighter program was used (<https://github.com/enjoping>)

References

- [1] Wikipedia, https://en.wikipedia.org/wiki/Alhazen's_problem
- [2] Dorrie Heinrich, "100 Great Problems of Elementary Mathematics, their history and solution", Dover Publications, New York 1965
- [3] Drexler, M., Gander, Martin Jakob. Circular Billiard. SIAM Review, 1998, vol. 40, no. 2, p.315 - 323 DOI: <https://doi.org/10.1137/S0036144596310872>
- [4] Elkin, Jack M, "A deceptively easy problem", The Mathematics Teacher, Vol 58, No 3(March 1985), pp 194,199
- [5] Miller, Allen R, & Vegh Emanuel, "Computing the grazing angle of specular reflection", Int.J.Math.Educ.Sci.Technol., 1990, Vol 21, No 2, pp271-274
- [6] Bach, Henning, "Some ray tracing problems related to circles and ellipses" , <https://apps.dtic.mil/sti/pdfs/ADA219306.pdf> , March 1989
- [7] Kollas Nikolaos, "Specular reflection on the surface of a sphere: compass and ruler constructions", <https://arxiv.org/pdf/1703.06768>

APPENDIX I: HOW FIND THE POINTS OF INTERSECTION OF TWO CONICS

Suppose we have two conics and we want to find their intersection points:

$$A_1 \cdot x^2 + B_1 \cdot x \cdot y + C_1 \cdot y^2 + D_1 \cdot x + E_1 \cdot y + F_1 = 0 \quad (1)$$

$$A_2 \cdot x^2 + B_2 \cdot x \cdot y + C_2 \cdot y^2 + D_2 \cdot x + E_2 \cdot y + F_2 = 0 \quad (2)$$

Multiplying (2) with $-C_2/C_1$ and adding it to (1) we get rid of the y^2 term and we get:

$$(A_1 \cdot C_2 - A_2 \cdot C_1) \cdot x^2 + (B_1 \cdot C_2 - B_2 \cdot C_1) \cdot x \cdot y + (D_1 \cdot C_2 - D_2 \cdot C_1) \cdot x + (E_1 \cdot C_2 - E_2 \cdot C_1) \cdot y + (F_1 \cdot C_2 - F_2 \cdot C_1) = 0 \quad (3)$$

Setting:

$$K = A_1 \cdot C_2 - A_2 \cdot C_1$$

$$M = D_1 \cdot C_2 - D_2 \cdot C_1$$

$$N = F_1 \cdot C_2 - F_2 \cdot C_1$$

$$Q = B_1 \cdot C_2 - B_2 \cdot C_1$$

$$P = E_1 \cdot C_2 - E_2 \cdot C_1$$

and plugging the values into (3) we have:

$$K \cdot x^2 + Q \cdot x \cdot y + M \cdot x + P \cdot y + N = 0$$

which can be solved for y as :

$$y = \frac{-K \cdot x^2 + M \cdot x + N}{Q \cdot x + P} \quad (4)$$

Plugging this into (3), after some algebra we get:

$$(A_1 \cdot x^2 + D_1 \cdot x + F_1) \cdot (Q \cdot x + P)^2 + C_1 \cdot (K \cdot x^2 + M \cdot x + N)^2 - (B_1 \cdot x + E_1) \cdot (Q \cdot x + P) \cdot (K \cdot x^2 + M \cdot x + N) = 0$$

and finally:

$$(A_1 \cdot Q^2 + C_1 \cdot K^2 - B_1 \cdot Q \cdot K) \cdot x^4 + \dots$$

$$(2 \cdot A_1 \cdot P \cdot Q + D_1 \cdot Q^2 + 2 \cdot C_1 \cdot K \cdot M - B_1 \cdot Q \cdot M - B_1 \cdot P \cdot K - E_1 \cdot Q \cdot K) \cdot x^3 + \dots$$

$$(A_1 \cdot P^2 + 2 \cdot D_1 \cdot Q \cdot P + F_1 \cdot Q^2 + C_1 \cdot M^2 + 2 \cdot C_1 \cdot K \cdot N - B_1 \cdot Q \cdot N - B_1 \cdot P \cdot M - E_1 \cdot Q \cdot M - E_1 \cdot P \cdot K) \cdot x^2 + \dots$$

$$(D_1 \cdot P^2 + 2 \cdot F_1 \cdot P \cdot Q + 2 \cdot C_1 \cdot M \cdot N - B_1 \cdot P \cdot N - E_1 \cdot Q \cdot N - E_1 \cdot P \cdot M) \cdot x + F_1 \cdot P^2 + C_1 \cdot N^2 - E_1 \cdot P \cdot N = 0 \quad (5)$$

This can be solved for x by a closed formula, and subsequently, y can easily be calculated from (4) for each real solution of (5)

ARITHMETIC EXAMPLE:

Conics:

conic1=(11.857699122508,12.256711089904,22.142300877492,-451.39059059008,-309.19000312823, 4146.705318439)

{Ellipse:a=5 * b=3 * Cx=18 * Cy=2 * tilt=155°}

conic2=(7.7432889100964,-20.569203509969,32.256711089904,-201.56421241337,156.13619313006,1335.0915158185)

{ Ellipse:a=6 * b=2 * Cx=17 * Cy=3 * tilt:20°}

Intermediate variables:

K=211.03614195703

M=-10097.280431974

N=104197.07738961

Q=850.81068146755

P=-13430.667168946

Coefficients of the quartic equation:

[x^4]=7368957.7320294

[x^3]=-496561171.96932

[x^2]=12406666090.275

[x^1]=-136290639530.17

[x^0]=555702343597.54

Solutions of the quartic equation (x) and corresponding values of y:

21.508863185604,3.1528212824536

17.589048643202,5.289277758176

14.98982518035,0.38298720057585

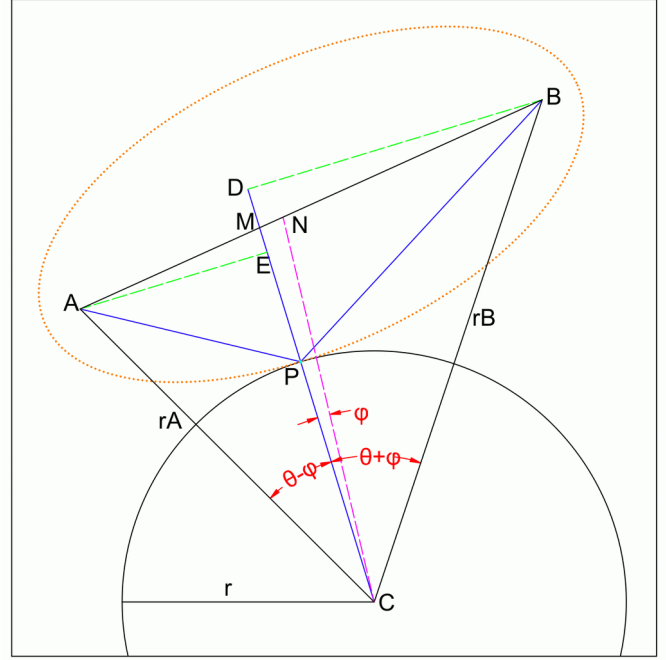
13.297799728422,3.4218839506969

APPENDIX II: ANOTHER CONFIGURATION TO SOLVE THE PROBLEM

An alternative way to solve Alhazen's problem is as follows:[7] The bisector CN of central angle ACB (see adjacent figure) forms an angle θ with each side. Suppose now that the reflection point is P. Then OP is the bisector of angle APC since the two rays form equal angles with the tangent at P. We draw perpendiculars from both A and B touching the bisector at points E and D respectively. The two orthogonal triangles AME and BMD are similar because they have the 3 angles equal. If ϕ is the angle between CN and OP we have:

$$\frac{r_A \cdot \sin(\theta - \phi)}{r_A \cdot \cos(\theta - \phi) - r} = \frac{r_B \cdot \sin(\theta + \phi)}{r_B \cdot \cos(\theta + \phi) - r} \quad (1)$$

where r_A, r_B are the segments CA and CB and r is the radius of the circle. So we have this nice equation to solve for ϕ . Once ϕ is found, P can be easily located by rotating CN by this angle. So, expanding (1) we get:



$$\frac{r_A \cdot [\sin(\theta) \cdot \cos(\phi) - \cos(\theta) \cdot \sin(\phi)]}{r_A \cdot [\cos(\theta) \cdot \cos(\phi) + \sin(\theta) \cdot \sin(\phi)] - r} = \frac{r_B \cdot [\sin(\theta) \cdot \cos(\phi) + \cos(\theta) \cdot \sin(\phi)]}{r_B \cdot [\cos(\theta) \cdot \cos(\phi) - \sin(\theta) \cdot \sin(\phi)] - r} \quad \text{or}$$

$$\frac{r_A \cdot \sin(\theta) \cdot \cos(\phi) - r_A \cdot \cos(\theta) \cdot \sin(\phi)}{r_A \cdot \cos(\theta) \cdot \cos(\phi) + r_A \cdot \sin(\theta) \cdot \sin(\phi) - r} = \frac{r_B \cdot \sin(\theta) \cdot \cos(\phi) + r_B \cdot \cos(\theta) \cdot \sin(\phi)}{r_B \cdot \cos(\theta) \cdot \cos(\phi) - r_B \cdot \sin(\theta) \cdot \sin(\phi) - r} \quad \text{or after some algebra}$$

$$-2 \cdot r_A \cdot r_B \cdot \sin(\phi) \cdot \cos(\phi) + r \cdot \sin(\theta) \cdot (r_B - r_A) \cdot \cos(\phi) + r \cdot \cos(\theta) \cdot (r_A + r_B) = 0$$

By replacing $\sin(\phi)$ with $\frac{2 \cdot \tan(\frac{\phi}{2})}{1 + \tan^2(\frac{\phi}{2})}$ and $\cos(\phi)$ with $\frac{1 - \tan^2(\frac{\phi}{2})}{1 + \tan^2(\frac{\phi}{2})}$ and setting $u = \tan(\frac{\phi}{2})$

we arrive at the following quartic equation for u :

$$r \cdot \sin(\theta) \cdot (r_B - r_A) \cdot u^4 - [4 \cdot r_A \cdot r_B + 2 \cdot r \cdot \cos(\theta) \cdot (r_A + r_B)] \cdot u^3 + [4 \cdot r_A \cdot r_B - 2 \cdot r \cdot \cos(\theta) \cdot (r_A + r_B)] \cdot u - r \cdot \sin(\theta) \cdot (r_B - r_A) = 0 \quad (2)$$

This equation can be solved for u with analytic methods and then ϕ can be calculated as $\phi = 2 \cdot \arctan(u \pm \pi)$

This equation seems to produce meaningful results no matter whether the points are located inside or outside the circle or they are collinear with a diameter, provided that A, B and C are distinct and r is greater than zero. The bisector of the angle ACB should be drawn from the center to the point that divides line segment AB internally into ratio AC/CB. If the angle between AC and bisector (as judged by their slopes) is greater than 90° the direction of bisector's rotation to find the reflection points should be positive otherwise negative, where positive means counterclockwise.

After each solution is found, the bisector of angle ACB must be rotated by that amount and its intersection points with the circle have to be calculated. Normally there are two such intersection points only one of which is the solution (or both in some symmetrical situations), resulting from the properties of $\arctan$ function. For each one of these points a check is run to see if the tangents to the circle and ellipse are the same.