How A Pencil under Water Looks:
Utilizing Astroidal Virtual Caustic for the Location of Refraction Image by Flat Boundary

M. Ryu

mingshey@hafs.hs.kr

February 8, 2025

1 Introduction

A pencil partially immersed in water, appearing bent, is a common phenomenon encountered when first learning about refraction (Figure 1). However, the apparent position of the pencil tip varies depending on the viewing angle, a phenomenon that warrants further investigation.
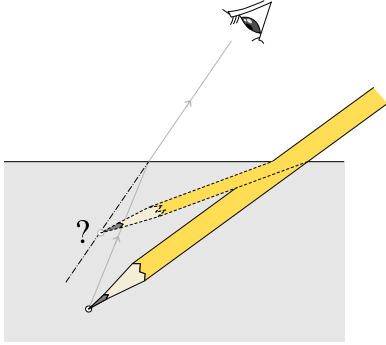

Figure 1: A pencil appearing bent in water

Typically, introductory physics courses cover the apparent depth when viewed vertically. However, a detailed analysis of how the apparent position changes when viewed obliquely is often omitted. Even advanced optics textbooks tend to gloss over this phenomenon, focusing instead on more fundamental topics such as lenses and mirrors.

This is perhaps because a rigorous analysis of this phenomenon requires a level of mathematical sophistication that may be beyond the scope of introductory courses, and its practical applications may be considered less significant compared to the workings of optical instruments.

However, even seemingly trivial phenomena can sometimes pique our curiosity. It is likely that many others have wondered about this phenomenon as well. This paper aims to provide an answer to this question.
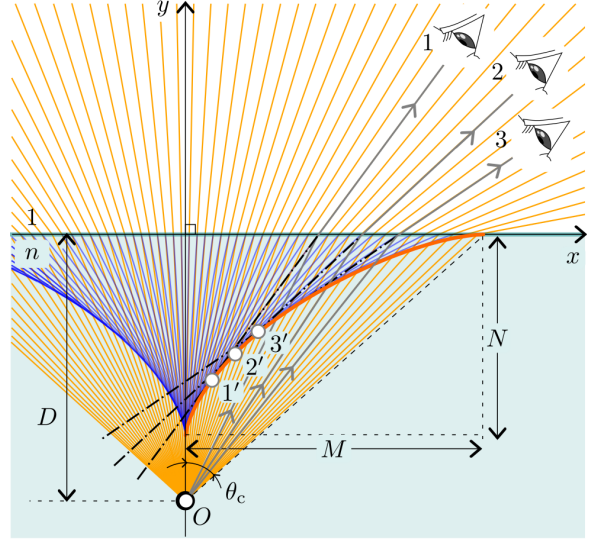

Figure 2: Caustic: the envelope of refracted rays extended

2 Quick Answer

For those who are curious but short on time, here's the bottom line: When viewing a point object submerged in water from above, the apparent position of the image varies depending on the point of view. As the viewing point moves, the locus of the observed image on the normal plane forms a type of caustic[1], which in this case is a curve known as a *squashed astroid*.

In the normal plane containing both the object and the viewer, let the intersection of the normal plane and the water surface be the x-axis, and the normal line passing through the object be the y-axis, as shown in Figure 2. Then, the locus of the image is part of the following curve:

$$\left|\frac{x}{M}\right|^{2/3} + \left|\frac{y}{N}\right|^{2/3} = 1,$$

where $M = D/\sqrt{n^2-1}$ is the maximum incident distance determined by the critical angle of total internal reflection (θ_c), $N = D/n$ is the apparent depth when

---

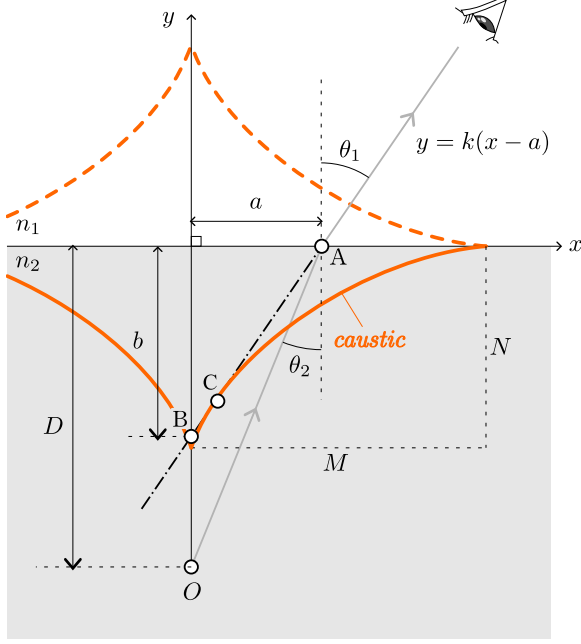[1] In this case you could call it a *virtual* caustic, since it's the locus of virtual images

Figure 3: Light rays and the caustic.

Figure 4: Projection of the caustic onto the $\xi\eta$ plane, revealing an astroid shape.

viewed directly from above, D is the actual depth of the object, and n is the refractive index of water relative to air.

3 Derivation of the Equation

Let n_1 and n_2 be the refractive indices of air and water, respectively. A point object O is located at a depth D below the air-water interface. A ray of light emanating from the object is incident on point A on the interface, which is a distance a from the y-axis, at an angle of θ_2 to the normal. The ray is then refracted into the air at an angle of θ_1 to the same normal.

The extension of the refracted ray intersects the y-axis at point $B(0, b)$. As the angle of incidence θ_2 varies, the lines AB envelope a curve, which is defined as the caustic curve. Note that in this case, the caustic is a virtual caustic formed by the extensions of the rays, rather than the rays themselves. The image is located at point C, the point of tangency between the line AB and the caustic. This is because the bundle of rays near this point diverges locally.

The straight line AB can be expressed by the following equation, where a is the distance from the y-axis to the point of incidence:

$$y = k(x - a).$$

Considering Snell's law, $n_1 \sin\theta_1 = n_2 \sin\theta_2$, or $\sin\theta_1 = n \sin\theta_2$, the slope of the line, k, is given by

$$k = \frac{1}{\tan\theta_1} = \frac{\cos\theta_1}{\sin\theta_1} = \frac{\sqrt{1 - n^2 \sin^2\theta_2}}{n \sin\theta_2}.$$

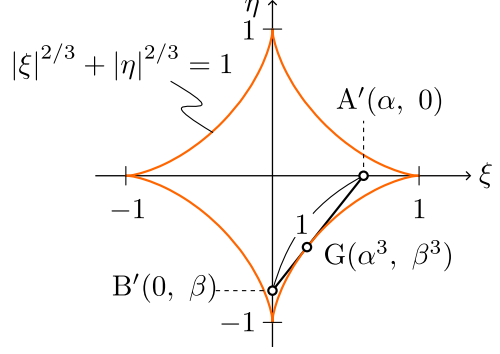Geometrically, the following relationships hold:

$$a = D\tan\theta_2 = \frac{D\sin\theta_2}{\cos\theta_2},$$

$$
\begin{aligned}
b &= -ka \\
&= -\frac{D\sin\theta_2}{\cos\theta_2} \cdot \frac{\sqrt{1 - n^2\sin^2\theta_2}}{n\sin\theta_2} \\
&= -\frac{D\sqrt{1 - n^2\sin^2\theta_2}}{n\cos\theta_2}.
\end{aligned}
$$

Introducing dimensionless parameters $\alpha = a/M$ and $\beta = b/N$, we have

$$
\begin{aligned}
\alpha^2 + \beta^2 &= \frac{a^2}{M^2} + \frac{b^2}{N^2} \\
&= \frac{n^2 - 1}{D^2}\frac{D^2\sin^2\theta_2}{\cos^2\theta_2} + \frac{n^2}{D^2}\frac{D^2(1 - n^2\sin^2\theta_2)}{n^2\cos^2\theta_2} \\
&= \frac{(n^2 - 1)\sin^2\theta_2 + 1 - n^2\sin^2\theta_2}{\cos^2\theta_2} \\
&= \frac{1 - \sin^2\theta_2}{\cos^2\theta_2} \\
&= 1
\end{aligned}
$$

Now, let's introduce dimensionless coordinates $\xi = x/M$ and $\eta = y/N$. As the viewpoint moves in the xy-plane, points $A(a, 0)$ and $B(0, b)$ also move. The corresponding points $A'(\alpha, 0)$ and $B'(0, \beta)$ move in the $\xi\eta$-plane maintaining a constant distance of 1.

The envelope of the family of lines $A'B'$ parameterized by α is a well-known curve called an *astroid*[2], which can be described by the equation:

$$|\xi|^{2/3} + |\eta|^{2/3} = 1.$$

(See Appendix A for details.)

The point of tangency between the line $A'B'$ and the astroid is $G(\alpha^3, \beta^3)$, which corresponds to the image point C.

---

[2]Not to be confused with the *asteroid* meaning a small celestial body.

Therefore, the coordinates of the image point (x_C, y_C) can be obtained from the following relations:

$$\begin{cases} \xi_G = \dfrac{x_C}{M} = \alpha^3 = \dfrac{a^3}{M^3}, \\[2mm] \eta_G = \dfrac{y_C}{N} = \beta^3 = \dfrac{b^3}{N^3}. \end{cases}$$

That is,

$$\begin{cases} x_C = \dfrac{a^3}{M^2}, \\[2mm] y_C = \dfrac{b^3}{N^2} = -\dfrac{k^3 a^3}{N^2}. \end{cases}$$

Using $\sin\theta_2 = a/\sqrt{D^2 + a^2}$ we obtain

$$k = \frac{\sqrt{D^2 - (n^2 - 1)a^2}}{na},$$

and the position of the image can be expressed as parametric functions of a:

$$\begin{cases} x_C = (n^2 - 1)\dfrac{a^3}{D^2}, \\[3mm] y_C = -\dfrac{D}{n}\left\{1 - (n^2 - 1)\dfrac{a^2}{D^2}\right\}^{3/2}. \end{cases}$$

4 When the viewpoint is submerged in water

When an object is in air at a height D above the boundary, and the viewer is in water, the relative refractive index is $1/n < 1$. By similar reasoning, we obtain the following equation for the caustic (see Appendix B):

$$|\xi|^{2/3} - |\eta|^{2/3} = -1,$$

where $\xi = x/W$ and $\eta = y/Z$, with $W = nD/\sqrt{n^2 - 1}$ and $Z = nD$. Although this curve does not have asymptotes, its slope converges to $\pm Z/W = \pm\sqrt{n^2 - 1}$ as $x \to \pm\infty$.

Thus, the skyscape viewed from underwater appears compressed within a circle (or cone) centered on the vertical axis and bounded by the critical angle for total internal reflection. This phenomenon is well-known as *Snell's window* and it gives field of view similar to that of a fisheye lens used in ultra wide-angle photography.

A more general form of this curve is the family of curves defined by

$$|\xi|^r - |\eta|^r = \pm 1,$$

which is referred to as *Super Hyperbola* by DML[3] and is included in a broader family of curves called *superconics* along with the family of curves called *superellipse* which is defined by

$$|\xi|^r + |\eta|^r = 1,$$

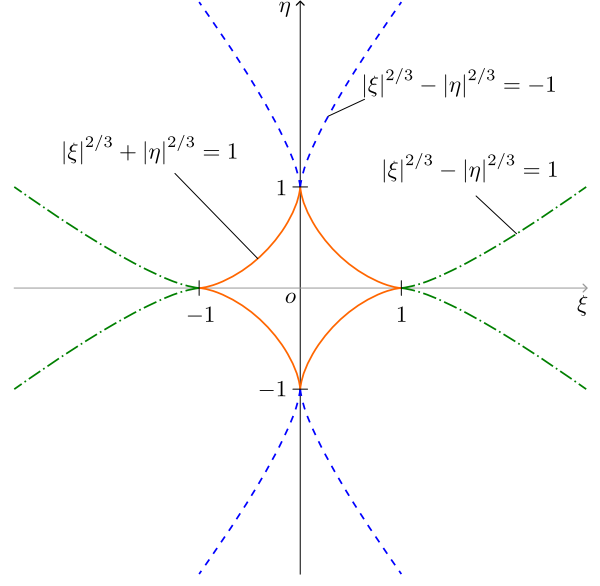[3]dynamicmathematicslearning.com/super-ellipse.html

Figure 5: Astroid and *'hyperastroids'*

as defined in the 'National Curve Bank'[4], to which astroid belongs.

In particular, for the special case where $r = 2/3$, the resulting curve

$$|\xi|^{2/3} - |\eta|^{2/3} = \pm 1,$$

has physical significance as the caustic of light rays emanating from a point source in the air and refracted into the water. Moreover, the relationship between this curve and the astroid is analogous to that between a hyperbola and an ellipse, suggesting the name *hyperastroid*.

5 Finding the Image Position

Given the position of an object and the viewpoint, the image position can be determined using the caustic as follows.

As shown in Figure 6, a tangent line is drawn from the viewpoint (POV) to the caustic. The point of tangency (O′) on the caustic is the image position, and the intersection point of the tangent line with the water surface is the point where the light ray from the object (O) enters the interface with air.

Let (x_V, y_V) be the coordinates of the viewpoint (POV), (x_O, y_O) of the object (O), and $(x_{O'}, y_{O'})$ of the image (O′) that we want to find. Let A and B be the points where the extension of the refracted ray intersects the water surface and the normal line passing through O (which is the y-axis), with coordinates $(a, 0)$ and $(0, b)$, respectively.

In the $\xi\eta$ coordinate system, the corresponding points are, the viewpoint: $(\xi_V, \eta_V) = (x_V/M, y_V/N)$,

---

[4]old.nationalcurvebank.org/superconicncb/
superconicncb.htm

Figure 6: Determining the image position using the caustic

the object: $(\xi_O, \eta_O) = (x_O/M, y_O/N)$, and the image: $(x_{O'}/M, y_{O'}/N) = (\alpha^3, \beta^3)$, while the coordinates of A′ and B′ are $(\alpha, 0) = (a/M, 0)$ and $(0, \beta) = (0, b/N)$, respectively. Here, $M = D/\sqrt{n^2 - 1}$, $N = D/n$, and $D = y_O$.

From the proportional relationship between the tangent, the water surface, and the η-axis, we obtain the following equation:

$$\frac{\eta_V}{\xi_V - \alpha} = -\frac{\beta}{\alpha}.$$

Using $\beta = \pm\sqrt{1 - \alpha^2}$ and squaring both sides, we obtain the following quartic equation for α:

$$(1 - \alpha^2)(\alpha - \xi_V)^2 = \alpha^2 \eta_V^2 \tag{1}$$

Generally, a quartic equation has four complex roots, and in this case, there are two real roots. The desired value of α is the one between 0 and ξ_V. Solving a quartic equation manually is generally difficult. Although an analytical solution can be obtained using a computer algebra system (CAS), it is often more efficient to find an approximate solution using numerical methods. (See Appendix C)

As can be seen from the form of the equation, the location of the image depends on that of the viewpoint. For example, as shown in Figure 7, the tip of a partially submerged pencil appears at different positions, 1′ and 2′, when viewed from points 1 and 2, respectively.

If there is a continuous object in water, as shown in Figure 8, we can find the corresponding image points $1', 2', 3', \dots$ for points $1, 2, 3, \dots$ on the object using the same method. The locus of these image points, as the points move continuously, forms the image of the object.

Another approach is to determine the intersection point of the ray with the water surface by solving the equation

$$\frac{n_1(x - x_V)}{\sqrt{y_V^2 + (x - x_V)^2}} + \frac{n_2(x - x_O)}{\sqrt{y_O^2 + (x - x_O)^2}} = 0,$$

Figure 7: Image of a pencil partially submerged in water. The image of the pencil tip moves from 1′ to 2′ as the viewpoint changes from 1 to 2.

Figure 8: Image of a continuous object. By connecting the image points $1', 2', 3', \dots$ corresponding to points $1, 2, 3, \dots$ on the object, the overall shape of the image can be drawn.

Figure 9: Photograph of a chopstick partially submerged in water. Taken from a low angle close to the water surface.

or after rearranging,

$$(x - x_V)^2 \left\{ y_O^2 + (x - x_O)^2 \right\} = n^2 (x - x_O)^2 \left\{ y_V^2 + (x - x_V)^2 \right\}, \tag{2}$$

which is obtained by applying Fermat's principle. The abscissa of the intersection point, which is a real root of the quartic equation lying between x_O and x_V, can then be used to determine the image position using the equation of the tangent point of the astroid. In fact, after further rearranging, equation (2) is equivalent to equation (1).

A Python example demonstrating how to find the image position using both analytical and numerical methods can be found at github[5]

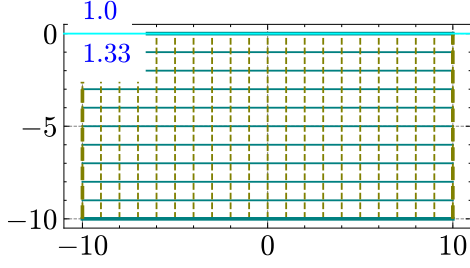[5] https://github.com/mingshey/python_projects/blob/main/Refraction_Image_en.ipynb

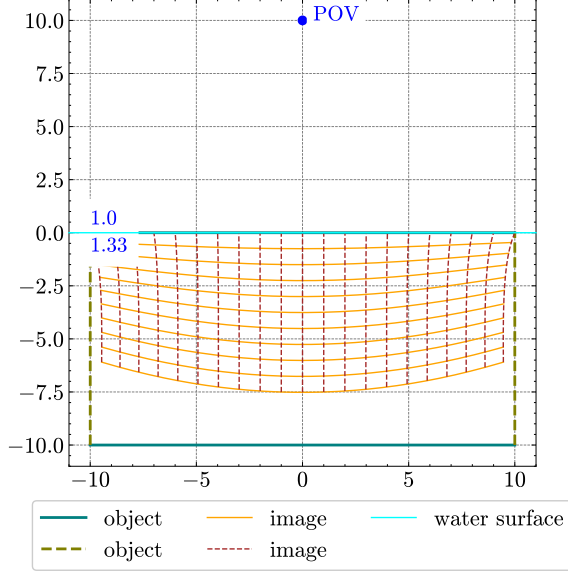**Figure 10:** A grid-shaped object placed underwater

Figure 11: Image of the grid. When the viewer's eye level is similar to or higher than the water depth, the image appears to be flattened overall in a ratio approximately equal to the refractive index.

The outline of the grid is shown as a rectangle for comparison in 11∼14

6 Image of the Submerged Space

Having established a quantitative method to locate the image, let's investigate how the overall appearance of the submerged space is distorted when viewed from above the water. As shown in Figure 10, a rectangular region in water represented by a square grid appears as shown in Figures 11 to 13 depending on the relative height of the viewpoint.

When a grid of equal squares is placed in the air adjacent to the water surface, the apparent shape of the grid viewed from underwater is compressed within a cone bounded by the critical angle of total internal reflection, as shown in Figure 14. The circle where the water surface intersects this cone is called Snell's window, and the entire sky above the water surface appears compressed within this window. The space directly above Snell's window is relatively undistorted and appears to be magnified by a factor equal to the refractive index ratio, while other regions appear distorted with compressed visual angles and increased
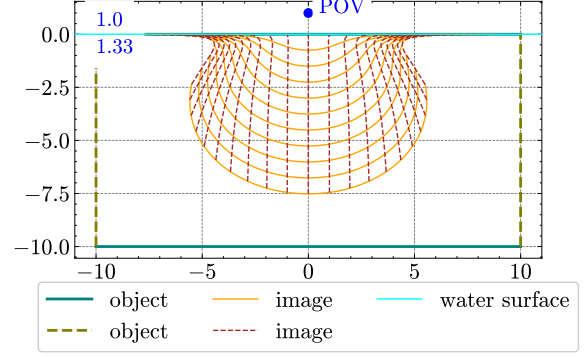

Figure 12: Image of the grid. When the viewer's eye level is much lower than the water depth, the rectangular space appears compressed into a fishbowl shape.
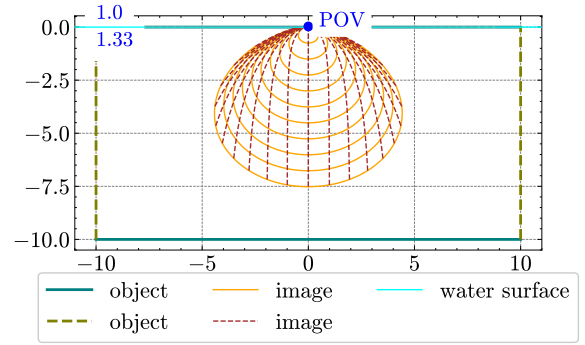

Figure 13: Image of the grid. When the viewer's eye is extremely close to the water surface, the rectangular space appears compressed into a shape resembling a seashell.
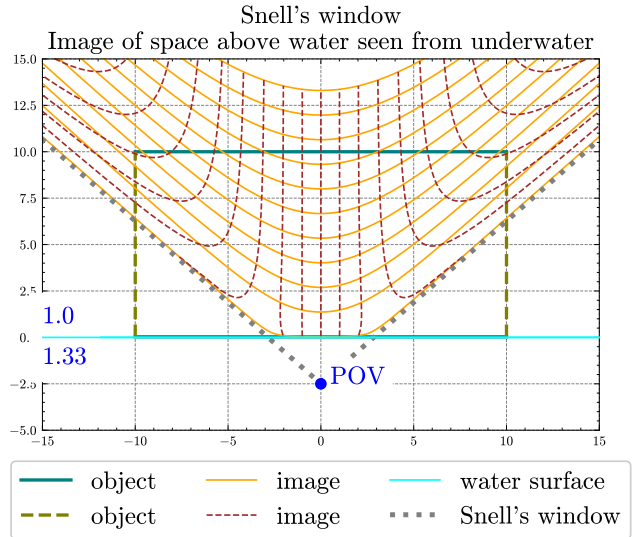

Figure 14: Snell's window. The appearance of an object viewed from underwater. The generating line of the cone defining Snell's window is conveniently denoted as Snell's window.

perceived distances. The distortion is particularly noticeable near the water surface.

Appendices

A The Astroid Defined as an Envelope

Consider two points in the Cartesian plane: $(a, 0)$ on the x-axis and $(0, b)$ on the y-axis. Suppose these points maintain a constant distance c from each other. Then, we have $a^2 + b^2 = c^2$, and the equation of the line connecting these two points at any given moment is given by

$$y = -\frac{b}{a}(x - a)$$

Substituting $b = \pm\sqrt{c^2 - a^2}$ yields

$$y(x, a) = \mp\frac{\sqrt{c^2 - a^2}}{a}(x - a)$$

As the value of a varies, the line connecting the two points also varies. The envelope of this family of lines can be defined as the locus of points that remain stationary as a undergoes an infinitesimal change, i.e., points where

$$\frac{\partial y}{\partial a} = 0.$$

Differentiating y with respect to a, we obtain

$$\frac{\partial y}{\partial a} = \pm\left[\left(\frac{1}{\sqrt{c^2 - a^2}} + \frac{\sqrt{c^2 - a^2}}{a^2}\right)(x - a) + \frac{\sqrt{c^2 - a^2}}{a}\right]$$

$$= \pm\frac{(a^2 + c^2 - a^2)(x - a) + a(c^2 - a^2)}{a^2\sqrt{c^2 - a^2}}$$

$$= \pm\frac{c^2 x - a^3}{a^2\sqrt{c^2 - a^2}}$$

$$= 0.$$

Therefore, the x-coordinate of the stationary point is $x = a^3/c^2$. Substituting this into $y(x, a)$, we obtain the y-coordinate:

$$y(x, a) = \mp\frac{\sqrt{c^2 - a^2}}{a}\left(\frac{a^3}{c^2} - a\right)$$

$$= \pm\frac{(c^2 - a^2)^{3/2}}{c^2}$$

$$= \frac{b^3}{c^2}$$

Therefore, the coordinates of the stationary point (x, y) satisfy the equation of the astroid:

$$\left|\frac{x}{c}\right|^{2/3} + \left|\frac{y}{c}\right|^{2/3} = 1.$$

B Hyperbolic Astroid

We define a 'hyperbolic astroid' as the envelope of the line segment connecting points $A(a, 0)$ on the x-axis and $B(0, b)$ on the y-axis, where these points satisfy the relation: $b^2 - a^2 = c^2$ The equation of this envelope is given by the following hyperbolic astroid equation:

$$\left|\frac{x}{c}\right|^{2/3} - \left|\frac{y}{c}\right|^{2/3} = -1.$$

The derivation is similar to that in Appendix A and is omitted here. If we draw a tangent line from a point A to the envelope, the coordinates of the tangent point are given by $(-a^3/c^2, b^3/c^2)$. If we draw a tangent line from an arbitrary point (x_V, y_V) to the hyperbolic astroid, and let $(a, 0)$ be the point where the tangent line intersects the x-axis, then a is a real root of the quartic equation

$$\left(a^2 + c^2\right)\left(a - x_V\right)^2 = a^2 y_V^2$$

lying between 0 and x_V, and $b = \pm\sqrt{a^2 + c^2}$.

C Python Code for Finding the Image Location Using the Tangent to the Astroid

```python
import numpy as np
import sympy as sym
from scipy.optimize import root

ksi, eta, alfa = sym.symbols('xi eta alpha')
eqn = (alfa - ksi)**2 * (1 - alfa**2)\
    - alfa**2 * eta**2
eqnf = sym.lambdify((ksi, eta, alfa), eqn);

def imgloc(pov, obj, nrel):
    xv, yv = pov
    xo, yo = obj
    M = yo / np.sqrt(nrel**2 - 1)
    N = yo / nrel
    xi_v = (xv - xo) / M
    eta_v = yv / N
    alpha = root(lambda x:\
        eqnf(xi_v, eta_v, x), xi_v).x[0]
    beta = np.sqrt(1 - alpha**2)

    xim = alpha**3 * M + xo
    yim = beta**3 * N
    xpoi = (xim*yv - yim*xv)/(yv - yim)

    img = np.array([xim, yim])
    poi = np.array([xpoi, 0])

    return img, poi
```

This document is translated from original Korean with the assistance of Google's Gemini language model.