Distance to an ellipse

Robert Estalella

2012 December

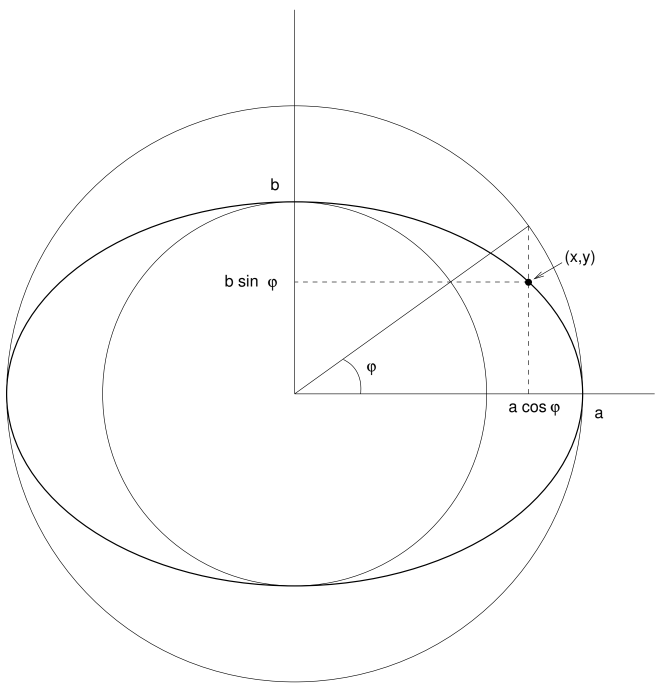
1 Distance of a point to an ellipse

Let us consider a centered ellipse with its axes along the x and y axes. The parametric equations of a point (x_e, y_e) of the ellipse are

$$\begin{aligned} x_e &= a\cos\varphi, \\ y_e &= b\sin\varphi, \end{aligned} \tag{1}$$

where a is the semimajor and b the seminimor axis, and φ is the angle defined in Fig. 1,

Figure 1: Geometry of the ellipse.



The distance l of a given point of the plane (x_p, y_p) to a point of the ellipse (x_e, y_e) is given by

$$l^2(\varphi) = (x_p - x_e)^2 + (y_p - y_e)^2 = (x_p - a\cos\varphi)^2 + (y_p - b\sin\varphi)^2, \tag{2}$$

which is a function of the angle φ. The distance of the point (x_p, y_p) to the ellipse is the minimum value of l, which occurs for a value of φ such that $d\,l^2/d\,\varphi = 0$. The derivative of the last equation gives

$$2(x_p - a\cos\varphi)a\sin\varphi - 2(y_p - b\sin\varphi)b\cos\varphi = 0, \tag{3}$$

which can be written as

$$x_p a\sin\varphi - y_p b\cos\varphi = (a^2 - b^2)\sin\varphi\cos\varphi. \tag{4}$$

This is an equation in φ that has to be solved to find the value of φ for which the distance is minimum. Once φ is found, the distance of the point (x_p, y_p) to the ellipse is given by

$$d = \sqrt{(x_p - a\cos\varphi)^2 + (y_p - b\sin\varphi)^2}. \tag{5}$$

2 Alternative derivation of Eq. 4

The equation in φ that gives the point of the ellipse nearest to the point (x_p, y_p) can be also derived from the fact that the straight line defined by the point (x_p, y_p) and the ellipse point (x_e, y_e) nearest to it is perpendicular to the ellipse at (x_e, y_e). Thus,

$$\frac{y_p - y_e}{x_p - y_e} = -\left(\frac{dx}{dy}\right)_e. \tag{6}$$

where $(dy/dx)_e$ is the slope of the ellipse at (x_e, y_e). From Eq. 1, we derive

$$\frac{dx}{dy} = \frac{dx/d\varphi}{dy/d\varphi} = \frac{-a\sin\varphi}{b\cos\varphi}, \tag{7}$$

and substituting into Eq. 6 we obtain

$$\frac{y_p - b\sin\varphi}{x_p - a\cos\varphi} = \frac{a\sin\varphi}{b\cos\varphi}, \tag{8}$$

which is the same equation as Eq. 4.

3 Solution of Eq. 4 through a 4th degree polinomial

Eq. 4 can be written as

$$\frac{x_p a}{\cos\varphi} - \frac{y_p b}{\sin\varphi} = c^2, \tag{9}$$

where $c^2 \equiv a^2 - b^2$. The change to the new variable $t = \tan(\varphi/2)$, with

$$\begin{aligned} \cos\varphi &= \frac{1-t^2}{1+t^2}, \\ \sin\varphi &= \frac{2t}{1+t^2}, \end{aligned} \tag{10}$$

transforms the last equation in φ into an incomplete 4th degree polynomial in t,

$$y_p b\, t^4 + (2x_p a + 2c^2)t^3 + (2x_p a - 2c^2)t - y_p b = 0. \tag{11}$$

Two of the four possible roots of the polynomial will give the points of the ellipse nearest to, and farthest from, the given point (x_p, y_p).

4 Iterative solution of Eq. 4

A simple iterative scheme to solve Eq. 4 is to write it in the form

$$\tan\varphi = \frac{(a^2 - b^2)\sin\varphi + y_p b}{x_p a}, \tag{12}$$

begin with an initial value for φ (for instance $\varphi_0 = 0$), and iterate

$$\varphi_{i+1} = \arctan\left[\frac{(a^2 - b^2)\sin\varphi_i + y_p b}{x_p a}\right]. \tag{13}$$

In order to avoid the indetermination in φ ($\tan\varphi = \tan(\varphi + \pi)$), it is best to consider that the distance from the points (x_p, y_p) and $(|x_p|, |y_p|)$ to the ellipse are the same, and modify slightly the last equation, to force that $0 \le \varphi \le \pi/2$,

$$\varphi_{i+1} = \arctan\left[\frac{(a^2 - b^2)\sin\varphi_i + |y_p| b}{|x_p| a}\right], \tag{14}$$

and calculate the distance as

$$d = \sqrt{(|x_p| - a\cos\varphi)^2 + (|y_p| - b\sin\varphi)^2}. \tag{15}$$

This iterative scheme has been tested to converge for a large sample of random ellipses and points. However, it could not converge for some pathological cases.

5 General case for a non-centered, non-aligned ellipse

Let us consider that, in general, the ellipse is centered on a point (x_0, y_0), and its major axis is at an angle θ from the x axis (θ growing from x to y). Let us call (x', y') the coordinates with origin at the ellipse center, and aligned along the ellipse axes. The coordinates of a point of the plane (x_p, y_p) in the new coordinate system are

$$\begin{aligned} x'_p &= (x_p - x_0)\cos\theta + (y_p - y_0)\sin\theta, \\ y'_p &= -(x_p - x_0)\sin\theta + (y_p - y_0)\cos\theta, \end{aligned} \tag{16}$$

while the inverse transformation is

$$\begin{aligned} x_p &= x_0 + x'_p\cos\theta - y'_p\sin\theta, \\ y_p &= y_0 + x'_p\sin\theta + y'_p\cos\theta. \end{aligned} \tag{17}$$

The distance of the point (x_p, y_p) to the non-centered, non-aligned ellipse is thus that of the point (x'_p, y'_p) to a centered and aligned ellipse, which can be computed as shown earlier.