Ellipse – Most Important Definitions and Facts

The ellipse is a special kind of **conic**. Unlike the other conics, it is a closed curve. It has a **centre** and two perpendicular **axes of symmetry**. The points of intersection of the axes with the ellipse are the **apeces** of the ellipse, which are also points of maximal/minimal curvature along the ellipse.

An ellipse with centre at the origin and apeces on the axes of a Cartesian coordinate system has an equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

or, in parametric form,

$$\begin{aligned} x &= a\cos t \\ y &= b\sin t, \end{aligned}$$

where a and b are the ellipse's **semimajor** and **semiminor axes** respectively. Such an ellipse is said to be represented in **canonical form**.

The ellipse has two focal points, or **foci**, symmetrically located around the centre, inside the ellipse and on the line connecting the two farthest apeces. The distance c from a focus to the centre is called the **focal distance**. (The distance between the foci is $2c$.) $c^2 = a^2 - b^2$ holds.

The ratio $\varepsilon = c/a \in [0, 1)$ is called **eccentricity** of the ellipse. When $\varepsilon = 0$, both foci coincide with the centre and the ellipse is a circle. The greater ε, the more oblate the ellipse.

An ellipse in canonical form has an **evolute** with the equation

$$(ax)^{2/3} + (by)^{2/3} = c^{4/3},$$

or, in parametric form,

$$\begin{aligned} x &= \frac{c^2}{a}\cos^3 t \\ y &= -\frac{c^2}{b}\sin^3 t. \end{aligned}$$

The evolute is an astroid with four apeces at $(\pm c^2/a, 0)$ and $(0, \pm c^2/b)$.

If a point is in the interior of the evolute, there are four normals from that point to the ellipse. From a point on the evolute but not an apex there are three normals. From any other point there are two normals.

For an ellipse in canonical form, denote the focus on the negative part of O_x as F_- and that on the positive part with F_+. For each point $P(x, y)$ on the ellipse, $r_- = \mathrm{dist}(F_-, P)$ and $r_+ = \mathrm{dist}(F_+, P)$ are called **focal radii**.

Remarkably, the focal radii are linear in x, namely: $r_- = a + \varepsilon x$ and $r_+ = a - \varepsilon x$ (r_- is increasing and r_+ is decreasing, both covering the range $[a-c, a+c]$). Moreover, the sum $r_- + r_+ = 2a$ does not depend on P.

If P is a point on the ellipse, the product of the distances of the foci to the tangent at P is a constant (does not depend on P).

For each point P on the ellipse, $\angle F_- P F_+$ is bisected by the normal to the ellipse at P.

Although an ellipse is not constructible with a ruler and compass, any individual point on it can be constructed. Here are two methods for doing this.

- Draw two concentric circles c_a with radius a and c_b with radius $b < a$. For each ray l from the centre, the point where the vertical line through the point of intersection of l and c_a and the horizontal line through the point of intersection of l and c_b meet is a point on the ellipse with semimajor axis a and semiminor axis b.

- Take two points P_1 and P_2 at a distance $2c$ from each other, and a point Q between them. The two points where the circle with centre P_1 and radius $\mathrm{dist}(P_1, Q) + a - c$ intersects with the circle with centre P_2 and radius $\mathrm{dist}(P_2, Q) + a - c$ are points on the ellipse with foci P_1 and P_2 and semimajor axis a ($a > c$).

The Ellipse As a Conic

Some important notions and facts of ellipses are meaningful and true of conics in general.

A conic is characterized as a set of points in the plane for which the ratio $\mathrm{dist}(P, F)/\mathrm{dist}(P, \Delta)$, where F is a given point and Δ is a given line, is constant. F is called the **focus** of the conic, Δ the **directrix**, and the ratio is the **eccentricity**. The value of the latter distinguishes one kind of conic from another: it is less than 1, equal to 1 and greater than 1 for an ellipse, parabola and hyperbola, respectively.

In case of the ellipse, the terms **focus** and **eccentricity** are exactly as already used. So for example ε describes the focal distance as a fraction of the semimajor axis, but also defines where the ellipse intersects the perpendicular from that focus to the directrix.

The directrix of an ellipse in a canonical form is parallel to O_y and at a distance $d = a/\varepsilon$ from O. In fact, for reasons of symmetry, there are two directrices, as there are two foci.

The chord of a conic through its focus that is parallel to the directrix is called the **focal chord**. The half-length of it, p, is called the **focal parameter** of the conic. p equals the minimal radius of curvature of the conic, the one at the crosspoint of the curve with the perpendicular from the focus to the directrix. This is b^2/a for an ellipse.

In a coordinate system with its origin in a focus and O_y parallel to the directrix of the conic, the equations of the conic are

$$x^2 + y^2 = (p - \varepsilon x)^2 \quad \text{(Cartesian)}$$

and

$$r = p/(1 + \varepsilon \cos t) \quad \text{(polar)}.$$

For an ellipse, the origin will be at F_+ and so r will correspond to r_+ (see above). From this equation, it is most straightforwardly seen that a circle is an ellipse with $\varepsilon = 0$.

Also, the equation

$$y^2 = 2px - (1 - \varepsilon^2)\, x^2$$

is sometimes used. It assumes that the coordinate system's origin is at a point on the conic: its apex. For the ellipse, this will be the nearest apex to F_-.

These equations express a conic in terms of ε and p. This is geometrically meaningful: the value of ε uniquely defines the conic up to a scaling factor. And with ε fixed, p is the scale factor of the curve. So ε defines the exact shape (including proportions) that characterize the conic, while p sets the scaling.

This can be seen on the ellipse: because $a = p/(1 - \varepsilon^2)$ and $b = p/\sqrt{1 - \varepsilon^2}$, both a and b are multiples of p, and thus changing p alters a and b in the same proportion.